Silesian University in Opava

School of Business Administration in Karviná

Czech Society of Operational Research

Czech Econometric Society

Proceedings of 30th International Conference Mathematical Methods in Economics

edited by

Jaroslav Ramík and Daniel Stavárek

Karviná, Czech Republic

11 – 13 September 2012

Editors: Jaroslav Ramík
Daniel Stavárek

Publisher: Silesian University in Opava, School of Business Administration in Karviná
Year of publishing: 2012
Number of pages: 1050, Part I. (pp. 1–523), Part II. (pp. 524–1050)

ISBN 978-80-7248-779-0

Suggested citation:
AUTHOR, A. Title of the paper. In: RAMÍK, J. and STAVÁREK, D. (eds.) *Proceedings of 30th International Conference Mathematical Methods in Economics.* Karviná: Silesian University, School of Business Administration, 2012, pp. xxx-xxx. ISBN 978-80-7248-779-0.

All published papers have been reviewed by two referees before publishing. The proceedings have not been amended or proofread and editors are not responsible for the language used in papers.

Programme Committee of the Conference:

Chair:

Prof. Dr. Jaroslav Ramík
Silesian University, School of Business Administration, Karviná, Czech Republic

Members:

Prof. Dr. Petr Fiala
University of Economics Prague, Faculty of Informatics and Statistics, Czech Republic

Assoc. Prof. Dr. Jana Hančlová
VŠB – Technical University Ostrava, Faculty of Economics, Czech Republic

Prof. Dr. Josef Jablonský
University of Economics Prague, Faculty of Informatics and Statistics, Czech Republic

Assoc. Prof. Dr. Ladislav Lukáš
University of West Bohemia Pilsen, Faculty of Economics, Czech Republic

Prof. Dr. Jan Pelikán
University of Economics Prague, Faculty of Informatics and Statistics, Czech Republic

Dr. Karel Sladký
Czech Academy of Sciences, Institute of Information Theory and Automation, Czech Republic

Assoc. Prof. Dr. Daniel Stavárek
Silesian University, School of Business Administration, Karviná, Czech Republic

Assoc. Prof. Dr. Tomáš Šubrt
Czech University of Life Sciences Prague, Faculty of Business and Economics, Czech Republic

Prof. Dr. Karel Zimmermann
Charles University, Faculty of Mathematics and Physics, Czech Republic

Organizing Committee of the Conference:

Conference website: **http://mme2012.opf.slu.cz**

Organizer's website: **http://www.slu.cz/opf/cz/**

List of Referees:

prof. RNDr. Jaroslav Ramík, CSc.
doc. Ing. Daniel Stavárek, Ph.D.
prof. Ing. RNDr. Petr Fiala, CSc.
prof. Ing. Josef Jablonský, CSc.
prof. Ing. Miloslav Vošvrda, CSc.
prof. Ing. Evžen Kočenda, Ph.D.
doc. Ing. Tomáš Šubrt, Ph.D.
doc. Mgr. Ing. Martin Dlouhý, Dr. MSc.
doc. RNDr. Helena Brožová, CSc.
doc. Ing. Jana Hančlová, CSc.
doc. RNDr. Jana Talašová, CSc.
doc. RNDr. Ing. Ladislav Lukáš, CSc.
prof. RNDr. Milan Vlach, DrSc.
prof. RNDr. Karel Zimmermann, DrSc.
prof. RNDr. Jan Pelikán, CSc.
prof. RNDr. Jaroslav Janáček, PhD.
prof. RNDr. Jitka Dupačová, DrSc.
prof. Ing. Václava Pánková, CSc.
prof. Ing. Roman Hušek, CSc.
prof. Ing. Jan Kodera, CSc.
prof. RNDr. Ing. František Turnovec, CSc.
prof. RNDr. Jan Hanousek, CSc.
prof. Ing. Dr. Zdeněk Zmeškal
prof. RNDr. Ján Plávka, PhD.
doc. RNDr. Petr Lachout, CSc.
Mgr. Jiří Mazurek, Ph.D.
Mgr. Radmila Stoklasová, Ph.D.
Ing. Pavla Vodová, Ph.D.
Ing. Irena Szarowská, Ph.D.
Ing. Zuzana Szkorupová, Ph.D.
Ing. Petr Seďa, Ph.D.
Ing. Jan Nevima, Ph.D.
Ing. Michal Tvrdoň, Ph.D.
Ing. Tomáš Tichý, Ph.D.
Ing. Jiří Valecký, Ph.D.
doc. Ing. Danuše Nerudová, Ph.D.
Ing. Karel Sladký, CSc.
Ing. Michal Kvasnička, Ph.D.
prof. RNDr. Jan Černý, DrSc.
Ing. Michal Černý, Ph.D.
Ing. Lenka Flusserová, Ph.D.
doc. RNDr. Jarmila Radová, Ph.D.
Ing. Jan Zouhar, Ph.D.
Ing. Jan Fabry, Ph.D.
Ing. Petr Rozmahel, Ph.D.
Ing. Ladislava Grochová, Ph.D.
doc. Ing. Svatopluk Kapounek, Ph.D.
Ing. Miroslav Čulík, Ph.D.
Ing. Rajmund Mirdala, PhD.
Mgr. Šárka Čemerková, Ph.D.
doc. Ing. Osvald Vašíček, CSc.
Ing. Radomír Perzina, Ph.D.
Ing. Filip Tošenovský, Ph.D.
Ing. Elena Mielcová, Ph.D.

CONTENTS

Part I.

Part II.

PREFACE

Dear conference participant,

It is a great pleasure to welcome you to the 30th MME 2012 Conference organized by the Czech Society of Operations Research, Czech Econometric Society in cooperation with the Silesian University in Opava, School of Business Administration in Karviná. The conference offers you the opportunity to meet the operations research and econometric community in the Czech Republic and also many researchers coming from Slovakia, Poland and other seven countries. It also offers you an exposure to the evolution of the several areas that compose Operational Research, Econometrics, and generally mathematical methods applied in economics to keep you updated on our continuously evolving dynamic disciplines. During the conference you will certainly meet old and new colleagues, exchange ideas, develop new projects. You will also feel and enjoy the special atmosphere of the eastern Ostrava-Karviná region of the Czech Republic.

Beside the scientific programme which includes more than 190 papers, the social programme is also very rich. Let me invite you particularly to the excursion offered by the organizers to Vítkovice Steel (part of the city of Ostrava), where production of iron, coal and agglomerates in so-called Bottom area (Dolní oblast) has been already terminated. Part of this strategic locality was proclaimed National Cultural Monument together with Hlubina coal mine. Here you can visit the original blast furnace, coal mine tower or the gas storage transformed into modern concert hall and other interesting sites.

This conference Proceedings is divided into two parts and includes 179 papers selected from more than 200 papers submitted to the conference programme committee. All published papers have been subjected to a strict reviewing procedure of two independent referees. A positive feature is that our Proceedings includes 33 papers published exclusively by young researchers – mostly doctoral students and more than 30 papers with young scientists as co-authors.

I am confident that you will find the 30th MME 2012 Conference stimulating, rewarding and pleasant and that you will enjoy your stay in Karviná.

In Karviná, September 2012

Prof. Dr. Jaroslav Ramík
Chair of the programme committee

Asset pricing in DSGE models — comparison of different approximation methods

Jan Acedański[1]

Abstract. There are many numerical methods suitable for approximating solutions of DSGE models. They differ in terms of accuracy, coding and computing time. However for many macroeconomic applications the differences in accuracy do not matter, since all methods generate approximations with similar statistical properties of simulated time series. In the paper we check whether this is also the case for DSGE models with financial variables, like stocks and risk-free bonds. These models are usually highly nonlinear and some special methods should be applied to approximate the asset prices. In the paper we take a simple macro-finance DSGE model proposed by Jermann, solve it with three different group of methods, simulate and check if the simulated series of financial variables differ in terms of basic statistical moments. For solving the model we use the higher-order perturbation approaches, the loglinear-lognormal method, as well as the Galerkin projection method. The results show that there might be significant differences between moments of the financial series in models approximated using different methods. For example for moderate parametrization the expected risk premium in the model approximated by the Galerkin method is about half of percentage point higher than for the perturbation methods and the loglinear lognormal approach. These results clearly indicate that the further research on the solution methods of DSGE models with financial variables is needed.

Keywords: DSGE models, asset pricing, solution methods, risk premium.

JEL classification: C63, C68, G12
AMS classification: 37H10

1 Introduction

Dynamic stochastic general equilibrium models (DSGE) are one of the main tools used for analysis of economic policy. Having solid microfoundations makes them from the one hand robust to the Lucas critique but also very complicated. From the mathematical point of view a model is represented by a set of stochastic, nonlinear difference equations. There have been many methods proposed in literature for approximation of such systems [7]. They differ in terms of accuracy, speed and implementation difficulty [1, 5, 6]. The most popular are perturbation methods based on local polynomial approximation of a solution, projection methods seeking for a global approximation with Chebyshev polynomials and approaches based on solving Bellman's optimality principle. Despite significant differences in terms of accuracy, as far as macroeconomic variables are concerned time series simulated from models approximated with different methods usually have similar statistical properties. Therefore for many applications the simplest first-order approximations, like loglinearisation, provide sufficient accuracy. However this may not be the case if a macroeconomic model is extended to include asset prices as well, since to price stocks or bonds correctly it is crucial to have correct second- and higher order moments of payoffs and discount factors.

In this paper we study statistical properties of asset prices in a DSGE model approximated using several different methods. Contrary to previous papers by Aruoba, Fernandez–Villaverde and Rubio–Ramirez [5] and Heer and Maussner [6], we use a modification of otherwise standard stochastic growth model proposed by Jermann [8] that incorporates exogenous habits in a utility function and investment costs. These modifications enable the model to generate a significant risk premium and therefore are commonly included in more complex macroeconomic models. In contrast to previous studies, the model

[1]University of Economics in Katowice, Department of Statistical and Mathematical Methods in Economics, 1-go Maja Street 50, 40-287 Katowice, Poland, jan.acedanski@ue.katowice.pl

exhibits higher nonlinearities, that make it particularly hard to approximate accurately. We solve the model using several perturbation methods of different orders, the Galerkin variant of the projection approach as well as the loglinear-lognormal approach which is a method tailored to approximating asset price dynamics in DSGE models. We show that although statistical properties of macroeconomic series in the model are virtually the same across the methods, the behaviour of asset prices may differ significantly which indicates the need for the further research in that area.

The paper is organized as follows. In the first section we briefly introduce the model. Then we discuss the methods used for the approximation. Finally we present the results of the simulation study.

2 The model

The paper uses the model proposed by Jermann [8] with only one minor modification — we abstract from the long-run growth. The economy is populated by large number of identical households who evaluate the consumption stream according to the instantaneous utility function:

$$u(C_t, C_{t-1}) = \frac{(C_t - \chi_c C_{t-1})^{1-\nu} - 1}{1-\nu}, \tag{1}$$

where C_t represents consumption, ν is the household's relative risk aversion, and χ_c is a habit persistence parameter. The households receive the income from work $W_t L_t$, where W_t is wage and L_t represents the fraction of time devoted to work. Since the households own firms, they also receive dividends D_t. So the budget constraint of the representative household has the following form:

$$W_t L_t + D_t = C_t. \tag{2}$$

In every period the household maximizes its expected lifetime utility:

$$\max_{C_t} \mathbb{E}_t \left[\sum_{h=0}^{\infty} \beta^h u(C_{t+h}, C_{t+h-1}) \right] \quad \text{s.t.} \quad C_{t+h} = W_{t+h} L_{t+h} + D_{t+h}, \quad h = 0, 1, ... \tag{3}$$

where β represents the household's discount factor.

The representative firm combines capital K_t with labour to produce single good Y_t according to a standard Cobb–Douglas technology:

$$Y_t = Z_t K_t^{\alpha} L_t^{1-\alpha}, \tag{4}$$

where α represents capital share in the output, whereas Z_t is a stochastic shock with AR(1) law of motion:

$$\ln Z_t = \rho \ln Z_{t-1} + \sigma \epsilon_t, \quad \epsilon_t \sim N(0,1). \tag{5}$$

The capital stock owned by the firm depreciates at a constant rate δ per period and is increased by investment I_t, so its evolution is given by:

$$K_t = K_{t-1} - \delta K_{t-1} + \Phi\left(\frac{I_t}{K_{t-1}}\right) K_{t-1}; \qquad \Phi\left(\frac{I_t}{K_{t-1}}\right) = \frac{a_1}{1-1/\xi}\left(\frac{I_t}{K_{t-1}}\right)^{1-1/\xi} + a_0. \tag{6}$$

Function Φ is a concave function capturing the idea that adjusting capital rapidly is more costly than changing it slowly. Each period the firm decides how much labour to hire and how much to invest trying to maximise utility of the dividend stream paid to the shareholders:

$$\max_{L_t, I_t, K_t} \mathbb{E}_t \left[\sum_{h=0}^{\infty} \beta^h M U_{t+h} D_{t+h} \right] \quad \text{s.t.} \quad K_{t+h} = \left[1 - \delta + \Phi\left(\frac{I_{t+h}}{K_{t+h-1}}\right)\right] K_{t+h-1}, \quad h = 0, 1, ..., \tag{7}$$

where a marginal utility of the household MU_t evolves according to:

$$MU_t = (C_t - \chi_c C_{t-1})^{-\nu} - \chi_c \beta \mathbb{E}_t \left[(C_{t+1} - \chi_c C_t)^{-\nu} \right] \tag{8}$$

and D_t corresponds to a net profit of the firm:

$$D_t = Y_t - W_t L_t - I_t \tag{9}$$

Since labour do not enter the utility function and its marginal product is always positive the households choose:

$$L_t = 1. \tag{10}$$

Wage equals the marginal product of labour:

$$W_t = (1-\alpha) Z_t K_{t-1}^{\alpha}. \tag{11}$$

The firm's first-order optimality conditions imply:

$$Q_t = \beta \mathbb{E}_t \left[\frac{MU_{t+1}}{MU_t} \left(\alpha Z_{t+1} K_t^{\alpha-1} - \frac{I_{t+1}}{K_t} + Q_{t+1} \left[1 - \delta + \Phi\left(\frac{I_{t+1}}{K_t} \right) \right] \right) \right], \tag{12}$$

where Q_t is a Lagrange multiplier associated with the capital law of motion constraint in the decision problem (7):

$$Q_t = \left[\Phi' \left(\frac{I_t}{K_{t-1}} \right) \right]^{-1}. \tag{13}$$

The model has 10 macroeconomic variables and consists of 10 equations: (2), (4)–(6), (8)–(13). It can also easily incorporate asset prices. For example, stock P_t and 1-period risk-free bond $P_{f,t}$ price dynamics are given by the standard formulas:

$$P_t = \beta \mathbb{E}_t \left[\frac{MU_{t+1}}{MU_t} (P_{t+1} + D_{t+1}) \right], \tag{14}$$

$$P_{f,t} = \beta \mathbb{E}_t \left[\frac{MU_{t+1}}{MU_t} \right]. \tag{15}$$

3 Approximation methods

The model introduced in the previous section can be compactly written as follows:

$$\mathbb{E}_t f(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \sigma \boldsymbol{\epsilon}_t) = \boldsymbol{0}, \quad \boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}), \tag{16}$$

where $\boldsymbol{X}_t$ is a vector of all n_x model variables, $\boldsymbol{\epsilon}_t$ represents a vector of n_e stochastic shocks and $f\colon \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_e} \to \mathbb{R}^{n_x}$. We look for a solution of the general form:

$$\boldsymbol{X}_t = g(\boldsymbol{X}_{t-1}, \sigma \boldsymbol{\epsilon}_t), \tag{17}$$

which can also be expressed in terms of a smaller set of state variables $\boldsymbol{S}_t$:

$$\boldsymbol{X}_t = g_x(\boldsymbol{S}_{t-1}, \sigma \boldsymbol{\epsilon}_t), \qquad \boldsymbol{S}_t = g_s(\boldsymbol{S}_{t-1}, \sigma \boldsymbol{\epsilon}_t). \tag{18}$$

For the analysed model the set of the state variables consists of consumption C_{t-1}, the capital stock K_{t-1} and the productivity shock Z_t. In the next subsections we discuss three approaches used for approximating the solution (17).

3.1 Perturbation methods

These methods are based on a local approximation of the solution with Taylor's polynomials around the model's deterministic steady state. This steady state can be seen as a limit of a system without any shocks, it is when $\sigma = 0$. Here we discuss only linearisation, the simplest of the perturbation approaches.

If we linearise the solution (17) around the steady state $\bar{\boldsymbol{X}}$, we get:

$$\boldsymbol{X}_t \approx \bar{\boldsymbol{X}} + \boldsymbol{G}_x (\boldsymbol{X}_{t-1} - \bar{\boldsymbol{X}}) + \boldsymbol{G}_\epsilon \sigma \boldsymbol{\epsilon}_t, \tag{19}$$

where $\boldsymbol{G}_x$ and $\boldsymbol{G}_\epsilon$ are Jacobians of the function g with respect to $\boldsymbol{X}_{t-1}$ and $\boldsymbol{\epsilon}_t$ respectively. The values of $\boldsymbol{G}_x$ and $\boldsymbol{G}_\epsilon$ can be calculated by inserting (17) into (16) and using the implicit function theorem, which states that since the l.h.s. of (16) is equal to 0, this must also be the case for all its derivatives. This

in turn leads to a matrix quadratic equation [6]. The linearised solution with the state variables has the following form:

$$\hat{\boldsymbol{S}}_t \approx \boldsymbol{M}\hat{\boldsymbol{S}}_{t-1} + \boldsymbol{W}\boldsymbol{\epsilon}_t, \qquad \hat{\boldsymbol{X}}_t \approx \boldsymbol{M}_X\hat{\boldsymbol{S}}_{t-1} + \boldsymbol{W}_X\boldsymbol{\epsilon}_t, \tag{20}$$

where hats over the variables represent deviations from the steady state: $\hat{\boldsymbol{S}}_t = \boldsymbol{S}_t - \bar{\boldsymbol{S}}$ and the matrices $\boldsymbol{M}, \boldsymbol{W}, \boldsymbol{M}_X$ and $\boldsymbol{W}_X$ consist of the elements of $\boldsymbol{G}_x$ and $\boldsymbol{G}_\epsilon$.

The higher-order approximations can be obtained in similar, recursive way by utilizing results of the approximation of a lower order as shown by Schmitt-Grohé and Uribe [9]. The method discussed here is very popular since it can be easily applied to models with large number of state variables. It is also very fast. Moreover it has been implemented in Matlab's Dynare package [3]. The main weakness of the approach is lack of accuracy since the obtained solution is close to the true one only near the steady state. When the system is far from the long-run equilibrium the approximation may be poor and in extreme cases the approximation may be divergent. However the higher-order approximations are considered by some researchers [5] to provide the accuracy of the similar order to other, more reliable approaches.

3.2 Loglinear lognormal approach

This method was proposed by Jermann [8] exclusively for the approximation of the asset price dynamics in linearised models. Since the linearisation abstracts from any second-order effects it cannot be used in models with asset prices. The loglinear-lognormal approach utilizes loglinear solution for the macroeconomic variables and extends it to include second-order terms for approximating asset prices. The loglinear solution of the model has the form (20):

$$\hat{\boldsymbol{s}}_t \approx \boldsymbol{M}\hat{\boldsymbol{s}}_{t-1} + \boldsymbol{W}\boldsymbol{\epsilon}_t, \qquad \hat{\boldsymbol{x}}_t \approx \boldsymbol{M}_X\hat{\boldsymbol{s}}_{t-1} + \boldsymbol{W}_X\boldsymbol{\epsilon}_t \tag{21}$$

where $\hat{\boldsymbol{s}}_t = \ln \boldsymbol{S}_t - \ln \bar{\boldsymbol{S}}$. From (15) we have:

$$\begin{aligned} P_{f,t} &= \beta\mathbb{E}_t\left[\frac{MU_{t+1}}{MU_t}\right] = \beta\mathbb{E}_t \exp\left[\hat{\lambda}_{t+1} - \hat{\lambda}_t\right] = \beta\mathbb{E}_t \exp\left[\boldsymbol{M}_\lambda\hat{\boldsymbol{s}}_t + \boldsymbol{W}_\lambda\boldsymbol{\epsilon}_{t+1} - \boldsymbol{M}_\lambda\hat{\boldsymbol{s}}_{t-1} - \boldsymbol{W}_\lambda\boldsymbol{\epsilon}_t\right] = \\ &= \beta\mathbb{E}_t \exp\left[\boldsymbol{M}_\lambda\left(\boldsymbol{M}\hat{\boldsymbol{s}}_{t-1} + \boldsymbol{W}\boldsymbol{\epsilon}_t\right) + \boldsymbol{W}_\lambda\boldsymbol{\epsilon}_{t+1} - \boldsymbol{M}_\lambda\hat{\boldsymbol{s}}_{t-1} - \boldsymbol{W}_\lambda\boldsymbol{\epsilon}_t\right] = \\ &= \beta\mathbb{E}_t \exp\left[(\boldsymbol{M}_\lambda\boldsymbol{M} - \boldsymbol{M}_\lambda)\hat{\boldsymbol{s}}_{t-1} + (\boldsymbol{W}_\lambda\boldsymbol{M} - \boldsymbol{W}_\lambda)\boldsymbol{\epsilon}_t + \boldsymbol{W}_\lambda\boldsymbol{\epsilon}_{t+1}\right] = \\ &= \beta \exp\left[(\boldsymbol{M}_\lambda\boldsymbol{M} - \boldsymbol{M}_\lambda)\hat{\boldsymbol{s}}_{t-1} + (\boldsymbol{W}_\lambda\boldsymbol{M} - \boldsymbol{W}_\lambda)\boldsymbol{\epsilon}_t + 0.5\boldsymbol{W}_\lambda'\boldsymbol{W}_\lambda\right], \end{aligned} \tag{22}$$

where in derivation we used the facts that: $\hat{\lambda}_t = \ln MU_t - \ln \bar{MU}$, $\hat{\lambda}_{t+1}$ and $\hat{\lambda}_t$ follow (21) and the expected value of the lognormal random variable $\mathbb{E}[\exp(x)] = \exp\left[\mathbb{E}(x) + 0.5\mathbb{D}^2(x)\right]$. The same approach can be applied to pricing stocks using discounted dividend version of the pricing equation, as shown by Jermann [8] and Acedański [2].

Similar to the perturbation approaches the presented method is easy to implement and can be applied to models with a large number of the state variables. However little is known about accuracy of the solution. Moreover the loglinear-lognormal framework treats the macroeconomic variables and the financial variables in different ways which sometimes is also considered as a weakness.

3.3 Projection method

The projection methods approximate globally either some parts of the solution (17) or some parts of the system (16) using linear combinations of Chebyshev polynomials $T_m(x)$ ($T_0(x) = 1$, $T_1(x) = x$, $T_m(x) = 2xT_{m-1}(x) - T_{m-2}(x)$, m – polynomial order). Following Heer and Maussner [6] we use both: we look for the approximating function $\Psi^{(1)}$ for the solution for Q_t and for the approximation $\Psi^{(2)}$ of the conditional expectation in the marginal utility dynamics (8) of the form:

$$\Psi^{(n)}(C, K, Z, \boldsymbol{\psi}^{(n)}) = \sum_{i,j,l} \psi^{(n)}_{i,j,l} T_i(C)T_j(K)T_l(Z), \quad i+j+l = m, \quad n = 1, 2, \tag{23}$$

For simplification we omit the time subscripts in the approximating functions (23). The unknown coefficients $\psi^{(1)}_{i,j,l}$ and $\psi^{(2)}_{i,j,l}$ should be set so to make the approximation functions as close as possible to the true functions within a given space of (C, K, Z). Then if we know the solution for Q_t we can easily find the conditional solution for all other variables. To find the values of $\psi^{(n)}_{i,j,l}$ distance measures $R^{(n)}$ need to

be defined. For the first equation $R^{(1)}(C, K, Z, \boldsymbol{\psi}^{(1)})$ is the difference between $\Psi^{(1)}$ and the r.h.s of (12) conditional on $\boldsymbol{\psi}^{(1)}$. For the second one $R^{(2)}(C, K, Z, \boldsymbol{\psi}^{(2)})$ is the difference between $\Psi^{(2)}$ and the expectation in (8) calculated conditional on $\boldsymbol{\psi}^{(2)}$. The expected values in both expressions are approximated using Gauss-Hermite quadrature formula with 10 nodes.

Then to find the values of $\psi^{(n)}_{i,j,l}$ we use the Galerkin condition: the coefficients should be set so to make the residuals orthogonal to the Chebyshev polynomials:

$$\int_{\underline{C}}^{\overline{C}} \int_{\underline{K}}^{\overline{K}} \int_{\underline{Z}}^{\overline{Z}} R^{(n)}(C, K, Z, \psi^{(n)}) \cdot T_i(C) \cdot T_j(K) \cdot T_l(Z)\, dZ\, dK\, dC = 0 \tag{24}$$

The quadratures above are approximated using Gauss-Chebyshev formula with 20 nods in each dimension. For both polynomials we use order $m = 6$, therefore the whole system of equations has 168 unknowns.

The projection methods are considered to be the most accurate [7], even far from the steady state. However they are hard to implement and are very slow. They also suffer from the curse of dimensionality. To approximate accurately the model with only 3 state variables the system of equations with more than 150 unknowns must be solved.

4 Results of the simulation study

For the simulation study we utilize rather standard parametrization and set $\alpha = 0.36$, $\beta = 0.99$, $\delta = 0.0136$, $\nu = 5$, $\rho = 0.95$, $\sigma = 0.01$. Only for the habit strength $\chi_c = 0.7$ and the curvature of the investment function $\xi = 0.8$ we use the values that slightly differ from the literature (Jermann [8] uses $\chi_c = 0.82$ and $\xi = 0.23$, whereas Heer and Maussner [6] study the model with $\chi_c = 0.8$ and $\xi = 0.23$). That parametrization do not allow the model to match the observed expected risk premium of about 6% per annum, but it makes the model less nonlinear and therefore easier to approximate, especially for the Galerkin method. We compare the following methods: perturbations in logs of the second-, third- and fifth-order, the loglinear-lognormal approach and the Galerkin method. For the given parametrization we approximate the model with these methods and run 1000 simulations with 250 quarters each.

Table 1 contains the basic moments of the macroeconomic variables. Since for all perturbation methods the results are exactly the same, we report them in one joint row. The table clearly shows that there are no important differences between the methods as far as the macroeconomic variables are concerned.

	Moments							
Method	$\mathbb{D}(Y)$	$\frac{\mathbb{D}(C)}{\mathbb{D}(Y)}$	$\frac{\mathbb{D}(I)}{\mathbb{D}(Y)}$	$\mathrm{ar}(Y)$	$\mathrm{ar}(C)$	$\mathrm{ar}(I)$	$\mathrm{corr}(C, Y)$	$\mathrm{corr}(I, Y)$
loglinear lognormal perturbations	0.013	0.52	3.25	0.71	0.90	0.57	0.87	0.95
Galerkin	0.013	0.57	3.19	0.71	0.90	0.55	0.87	0.95

HP-filtered quarterly averages over 1000 simulations of 250 quarters; ar – autocorrelation coefficient; corr – correlation coefficient.

Table 1 Moments of the simulated macroeconomic variables

In table 2 we report the moments of the financial variables. Two observations are worth noting. First, as far as the perturbations and loglinear-lognormal approach are concerned the differences between the moments are negligible but still at least of one order of magnitude higher than in case of the macroeconomic variables. And second, there are significant differences between the moments for the Galerkin approach and the rest of the methods, especially in case of the expected risk premium. For the former the premium is about 1.4 percentage point, whereas for the latter it is less than 1 percentage point. So there is more than 40% difference, which for more extreme calibrations can be much higher in absolute values. The nonnegligible differences are also observed for the standard deviation of the risk-free rate (5.4 p.p. – 4.7 p.p.), the standard deviation of the dividend growth rate (2.9 p.p – 3.15 p.p) and the expected dividend/price ratio (4.3 – 4.0). These results are in sharp contrast with Jermann [8] who found no differences between the projection method and loglinear-lognormal approach in his model, but are supported by the results of Aldrich and Kung [4], who also obtained significant differences in asset price moments for the projection method and the standard perturbation techniques.

Method	Moments								
	$\mathbb{E}(R)$	$\mathbb{E}(R_f)$	$\mathbb{E}(R-R_f)$	$\mathbb{D}(R)$	$\mathbb{D}(R_f)$	$\mathbb{D}(\Delta D)$	$\mathbb{E}(DP)$	$\mathbb{D}(DP)$	$\mathrm{corr}(R, R_f)$
loglin-lognor	4.44	3.54	0.90	8.34	4.70	3.12	4.04	0.50	-0.50
perturb. 2	4.39	3.43	0.96	8.35	4.65	3.16	3.99	0.49	-0.50
perturb. 3	4.43	3.57	0.86	8.34	4.71	3.17	4.02	0.50	-0.50
perturb. 5	4.41	3.47	0.94	8.34	4.68	3.17	4.00	0.49	-0.50
Galerkin	4.71	3.31	1.40	8.27	5.40	2.88	4.32	0.52	-0.49

Annualized averages in p.p. over 1000 simulations of 250 quarters; $\mathbb{E}$ – unconditional mean, $\mathbb{D}$ – unconditional standard deviation, corr – correlation coefficient, R – stock return, R_f – risk-free rate, ΔD – dividend growth rate, DP – dividend/price ratio.

Table 2 Moments of the simulated financial variables

5 Conclusion

In the paper using the model proposed by Jermann we have shown that despite the differences in accuracy all the approximation methods generate virtually the same moments of the main macroeconomic variables. However for the financial variables the differences between the perturbation method and the projection approaches are much higher. For moderate parametrization the expected risk premium in the model approximated by the Galerkin method is about 0.5 percentage point higher than for the perturbation methods and the loglinear lognormal approach. But it must be made clear that although the projection methods are considered to be the most accurate we cannot find out which method gives the moments that are closer to the true values in that particular case. Nonetheless these results clearly indicate that the further research on the solution methods of DSGE models with financial variables is needed.

References

[1] Acedański, J.: Solution methods of dynamic, stochastic general equilibrium models. In: *Contemporary problems of statistics, econometrics and mathematics* (J. Pociecha, ed.), University of Economics in Kraków, Kraków, 2009, 7–19 (in Polish).

[2] Acedański, J.: *Asset prices and macroeconomic variables in dynamic, stochastic general equilibrium models*. PhD Dissertation, University of Economics in Katowice, 2010 (in Polish).

[3] 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.

[4] Aldrich, E., and Kung, H.: Computational methods for production-based asset pricing models with recursive utility. Mimeo.

[5] Aruoba, S., Fernandez–Villaverde, J., and Rubio–Ramirez, J.: Comparing solution methods for dynamic equilibrium economies. *Journal of Economic Dynamics and Control* **30** (2006), 2477–2508.

[6] Heer, B., and Maussner, A.: *Dynamic General Equilibrium Modelling. 2nd edition*. Springer, Berlin, 2008.

[7] Heer, B., and Maussner, A.: Computation of business cycle models: A comparison of numerical methods. *Macroeconomic Dynamics* **12** (2008), 641–663.

[8] Jermann, U.: Asset pricing in production economies. *Journal of Monetary Economics* **41** (1998), 257–275.

[9] Schmitt-Grohé, S., and Uribe, M.: Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal of Economic Dynamics and Control* **28** (2004), 755–775.

Rule-of-thumb households in the Czech Republic

Tomáš Adam[1], Jaromír Baxa[2]

Abstract. Most modern macroeconomic models assume that households smooth consumption over their lifetime. However, there is substantial evidence that a sizable fraction of households faces liquidity constraints, thus the assumption of consumption smoothing is not met. This might be one of the reasons why the predictions of some models fail, for example those concerning effects of fiscal policy on consumption. To overcome this problem, some models assume that only a part of households (Ricardian households) smooth their consumption and the other part (rule-of-thumb households) consume their whole current disposable income. This paper estimates the fraction of rule-of-thumb households in the Czech Republic, based on the modified Euler equation using instrumental variables (approach by Campbell and Mankiw [4]). Furthermore, potential time variation in this parameter is investigated using the two-step strategy proposed by Kim [10] that deals with the endogeneity problem. Our results suggest that the share of rule-of-thumb consumers exceeds 40%. In addition, this share is unstable over time, however the decline in the recent years is puzzling and worth exploring further.

Keywords: consumption, rule-of-thumb households.

JEL classification: E21
AMS classification: 91G70

1 Introduction

The current mainstream macroeconomic models based on microfoundations used for the analysis of economic policies and for forecasting (RBC and NK DSGE models) use consumption smoothing as one of their main building blocks. This follows from the assumption that households maximize their lifetime utility subject to their intertemporal budget constraints. However, it has been shown that the assumption of consumption smoothing is not met in reality because some households face liquidity constraints or behave in a myopic way. One way to model the departure from consumption smoothing is to assume that a fraction of consumers in the economy exhibit a rule-of-thumb behaviour, i.e. they consume their whole current income instead of their permanent income, as suggested by the theory. Although this is still a relatively strong assumption, models incorporating rule-of-thumb households give usually more plausible results than those assuming fully optimizing behaviour only (e.g. [7]).

This is particularly the case of models for analysing fiscal policy. It has been shown that both RBC and New Keynesian models fail to predict the behaviour of consumption following a government spending shock (e.g. [11]), which has been empirically shown to rise. Incorporating rule-of-thumb households mitigates this problem. For example, Galí et al. [7] show that fiscal expansion that increases incomes of these rule-of-thumb households will have a direct effect on their consumption, since it is directly influenced by the current income and not by net present value of all future incomes. Furthermore, when the share of the rule-of-thumb households is large, fiscal policy affecting the incomes of rule-of-thumb households has direct macroeconomic effects on the aggregate consumption. Therefore, the consumption might increase in response to an increase in government expenditures, which is in a sharp contrast to the predictions given by most of the current RBC and NK DSGE models. Similarly, the presence of rule-of-thumb households has policy implications for fiscal consolidations: their negative effects on economic

[1]Charles University in Prague, Faculty of Social Sciences, Institute of Economic Studies, Opletalova 26, 110 00, Praha 1, e-mail: tomas.adam@ies-prague.org

[2]Institute of Information Theory and Automation/Academy of Sciences of the Czech Republic, Department of Econometrics, Pod Vodarenskou vezi 4, 182 08 Prague 8; Charles University in Prague, Faculty of Social Sciences, Institute of Economic Studies, Opletalova 26, 110 00, Praha 1, e-mail: jaromir.baxa@gmail.com

growth tend to be more pronounced in economies, where consumption of an important share of households follows the rule-of-thumb pattern rather than consumption smoothing.

In this paper, we estimate the share of rule-of-thumb households in the Czech Republic using a two stage least squares regression suggested in [4]. We show that this fraction has been significant, which suggests that rule-of-thumb behaviour should be taken into account in macroeconomic models. In addition, we attempt to estimate time-varying share of these households. Our results are however puzzling in that they show some decline before 2007 but a sharp drop after that, which is not expected. This issue is subject to future research.

2 Related literature

The life cycle hypothesis and the permanent income hypothesis ([3], [6]) imply that consumers smooth their consumption over their lifetimes. Both approaches are in a sharp contrast to the theory of consumption by Keynes [9] who claims that the level of current consumption is a function of the current disposable income only. Instead, the two more recent theories assert that consumers optimize their consumption profiles intertemporally through their lifetime, taking into account their income profiles. In their optimization, consumers regard their consumption in each period as a different commodity and maximize their lifetime utility, which provides the theory of consumption with solid microfoundations. The predictions of both, permanent income and life cycle hypotheses, are very similar - the optimization of a concave utility function leads to consumption smoothing, i.e. one does not want her consumption to fluctuate through time.

The permanent income hypothesis has been widely empirically tested with mixed results ([1] is a good starting point for a literature survey). For example, Hall [8] finds an evidence for the modified version of the hypothesis – i.e. that consumption follows a random walk, if we assume that changing the volume of consumption is time demanding. That implies that any policy that does not affect permanent income is inefficient. On the other hand, Campbell and Mankiw [4] claim that permanent income hypothesis can be rejected. This is because there are two types of consumers – one that behaves optimally and smooths its consumption over lifetime; the other type consumes its whole current income (due to liquidity constraints, myopia or unwillingness to participate in financial markets). The share of the second type of households was estimated as highly significant and reached about 0.5 in the USA under various specifications. This means, for example, that additional, even transitory, income will be spent by the second type of households and thus fiscal spending is a plausible means to stimulate output. In his additional paper, Mankiw [12] summarizes the evidence of rule-of-thumb behaviour, which characterizes particularly households, whose net wealth approaches zero. As this behaviour is common in the economy, he calls for the inclusion of rule-of-thumb households into all models analyzing macroeconomic, particularly fiscal, policies.

3 Methodology

The share of rule-of-thumb households in an economy is usually estimated using two approaches. The first one was introduced by Campbell and Mankiw [4] as a means of testing the permanent income hypothesis. The second approach estimates this share as a parameter of a DSGE model (e.g. [5]). The drawback of the latter approach is that the estimation is sensitive to the specification of the model and priors for the model's parameters, thus we will use the first approach.

In can be easily shown that the permanent income hypothesis implies that under some weak assumptions, consumption follows a random walk. If a fraction λ of households is assumed to consume their whole current disposable income, we arrive at the following modified Euler equation (e.g. [5]):

$$\Delta c_t = \alpha + \lambda \Delta y_t + \epsilon_t \tag{1}$$

where c_t is personal consumption and y_t is personal disposable income. Under the permanent income hypothesis, λ should be equal to zero, i.e. change in consumption should follow a random walk.

3.1 Time invariant estimates

The estimation of the time invariant share of rule-of-thumb households is relatively straightforward and follows from Equation 1. However, one must bear in mind that this equation is endogenous (ϵ_t is correlated with Δy_t) and thus an instrumental variable approach needs to be used. As Campbell and Mankiw [4] suggest, any lagged variables that help to predict changes in income can be used, since they are uncorrelated with changes in consumption.

3.2 Time-varying estimates

It is plausible that λ is not constant over time ([2], [13]). This is particularly true for the case of the Czech Republic, which has undergone a process of economic transformation, during which the financial system has changed significantly. Various credit institutions have been set up, earnings have become more dispersed among the population, both of which have led to changing structure of liquidity constraints and rule-of-thumb behaviour.

The time-varying λ can be estimated in the framework of time-varying regression with endogenous covariates. A standard Kalman filter approach cannot be used, so we use a method by Kim [10] to estimate the share of rule-of-thumb households.

We assume that the parameters in the endogenous regression follow a random walk:

$$\begin{cases} \Delta c_t = \alpha_t + \lambda_t \Delta y_t + e_t & \\ \alpha_t = \alpha_{t-1} + u_{\alpha,t} & u_{\alpha,t} \sim N(0, \sigma_\alpha^2) \\ \lambda_t = \lambda_{t-1} + u_{\lambda,t} & u_{\lambda,t} \sim N(0, \sigma_\lambda^2) \end{cases} \tag{2}$$

Similarly, parameters in the first stage are also assumed to follow random walks:

$$\begin{cases} \Delta y_t = z_t' \delta_t + \sigma_v v_t^* & \\ \delta_t = \delta_{t-1} + u_{\delta,t} & u_{\delta,t} \sim N(0, \sigma_{delta}^2) \end{cases} \tag{3}$$

The endogeneity in the regression is assumed to have the following form:

$$\begin{pmatrix} v_t^* \\ e_t \end{pmatrix} \sim iidN \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_e \\ \sigma_e\rho & \sigma_e^2 \end{pmatrix} \right) \tag{4}$$

Then a relatively straightforward procedure can be used (see [10] for its derivation):

- Step 1: Estimate a time-varying equation $\Delta_t = z_t' \delta_t + \sigma_v v_t^*$ using Kalman filter, obtain $v_{t|T}^*$
- Step 2: Estimate an adjusted time-varying equation: $\delta c_t = \alpha_t + \lambda_t \Delta Y_t + \sigma_e \rho v_{t|T}^* + \omega_t$

4 Data

Two main variables - consumption and income - in Equation 1 had to be chosen. We use consumption of non-durable goods and services as a variable representing consumption. This variable has been downloaded from the Czech Statistical Office. Next, we use GDP as the variable for income. GDP is only a proxy and it has been chosen because the variable on disposable income has been constructed only since 2007 by the CZSO[1].

The time series are deflated by consumption / GDP deflator and are transformed to logarithmic per capita values. Lagged changes in the two variables, along with changes in 3 month money market rates are used as instruments. In addition, we use an error correction term $c_{t-2} - y_{t-2}$ of log real per capita values as an instrument.

[1] A similar variable was published by the Czech National Bank but it is not constructed any more. Unfortunately, the methodology for the two differs and they are not comparable

5 Results

5.1 Time invariant share

Several instruments have been used to account for endogeneity in Equation 1: lagged changes in real per capita consumption, income, interest rates and finally an error correction term $c_{t-2} - y_{t-2}$. As Campbell and Mankiw [4] argue, variables at least at two lags should be used to prevent autocorrelation in the results, thus we follow this advice.

Based on R^2 from the first stage, the models reported in Table 1 are used for the estimation. The reported data in Table 1 are estimated on the sample ending in 2007Q4. This is because the development of GDP growth in the period of financial crisis was hardly predictable using macroeconomic data and including the mentioned data could distort our results.

Model	Instruments	n	R^2	F
IV(1)	$\Delta y_{t-2} \dots \Delta y_{t-4}$	43	0.52	15.6
IV(2)	$\Delta y_{t-2} \dots \Delta y_{t-6}$	41	0.47	7.5
IV(3)	$\Delta y_{t-2} \dots \Delta y_{t-4}, \Delta c_{t-2} \dots \Delta c_{t-4}, c_{t-2} - y_{t-2}$	43	0.53	7.88
IV(4)	$\Delta y_{t-2} \dots \Delta y_{t-4}, \Delta c_{t-2} \dots \Delta c_{t-4}, \Delta i_{t-2} \dots \Delta i_{t-4}, c_{t-2} - y_{t-2}$	43	0.55	6.2

Table 1: Adjusted R^2 and F statistics from the first stage

The second stage results reported in Table 2 suggest that the share of rule-of-thumb households has been significant during the period under consideration. The estimates vary among the four considered models and range from 0.38 to 0.59. This is in line with the results for other countries. For example, [4] have estimated this share to be 0.5 for the US and 0.4 in Italy. Similarly, [5] estimate this share to be 0.37 in their DSGE model of the eurozone. Also, a dummy variable indicating the financial crisis period was added as an explanatory variable but the results have not changed much (thus they are not reported in the table).

	OLS	n	IV(1)	IV(2)	IV(3)	IV(4)
1996Q1 - 2011Q4	0.24	63	0.59	0.49	0.38	0.43
1996Q1 - 2007Q4	0.41	47	0.85	0.7	0.81	0.81
1998Q1 - 2011Q4	0.24	56	0.41	0.37	0.34	0.33
1998Q1 - 2007Q4	0.44	40	0.39	0.15	0.41	0.32

Table 2: Estimated share of rule-of-thumb households (λ) using 2SLS

In contrast, the results vary when the models are estimated on three subsamples - the pre-crisis period, the period starting in 1998Q1 and the period 1998Q1-2007Q4[2]. The varying results point to the instability of the share and suggest that the share has fallen after 2008, which is a puzzling result (one would expect the share to increase due to deteriorating credit conditions) worth further investigation. The first explanation might be that the fall in income was expected to be only short-lived, thus the households did not consume so much less as implied by the change in income (this would be in line with the permanent income hypothesis). The second explanation might be a misspecification of the model or problems connected with the definition of the variables. Both possibilities will be explored in our further research.

5.2 Time-varying share

The puzzling result from the previous section, i.e. the fall in the share, is confirmed also using the time-varying model estimated using the approach by [10]. The results tend to be in line with the permanent income hypothesis since around 2007. This is in contrast with the evidence, i.e. the deteriorating credit conditions of households.

[2]Statistical significance of instability of the estimates was confirmed also by CUSUM and CUSUM-SQ tests.

Figure 1: Time-varying share of rule-of-thumb households in the Czech Republic (1997Q2 - 2011Q4)

6 Conclusion

In this paper, we have discussed the importance of incorporating rule-of-thumb behaviour of consumers into macroeconomic models and we have performed both time-invariant and time-varying estimation of the share of rule-of-thumb households. Our results suggest that the permanent income hypothesis is rejected in the Czech economy due to a significant share of rule-of-thumb households. However, the results point to a decline in the recent years, which is not a very plausible result due to deteriorating credit conditions of households. In our future work, we will try to find factors that are behind this decline and find a credible robust estimate of the share that could be used in macroeconomic models.

Acknowledgements

The support from the Grant Agency of Charles University in Prague (GAUK, project number 641012), is gratefully acknowledged.

References

[1] Attanasio, O.: Consumption, *Handbook of macroeconomics*, **1**, 1999, 731–812.

[2] Bacchetta, P. & Garlach, S.: Consumption and credit constraints: International evidence, *Journal of Monetary Economics* **40** (1997), 207–238.

[3] Brumberg, R. & Modigliani, F.: Utility analysis and the consumption function: An interpretation of cross-sectional data, *Post Keynesian economics* (1954).

[4] Campbell, J. and Mankiw, N.: Consumption, income, and interest rates: Reinterpreting the time series evidence, *NBER macroeconomics annual* (1989), 185–216.

[5] Coenen, G. & Straub, R.: Non-Ricardian households and fiscal policy in an estimated DSGE model of the euro area, *Manuscript, ECB* (2004).

[6] Friedman, M.: *A theory of the consumption function*. National Bureau of Economic Research, 1957.

[7] Galí, J. and López-Salido, J.D. and Vallés, J.: Understanding the effects of government spending on consumption, *Journal of the European Economic Association* **5** (2007), 227–270.

[8] Hall, R.: Stochastic implications of the lie cycle-permanent income hypothesis: Theory and evidence, *The Journal of Political Economy* **86** (1978), 971–987.

[9] Keynes, J. M.: *The General Theory of Employment, Interest and Money*, London: MacMillan, 1936.

[10] Kim, C.: Time-varying parameter models with endogenous regressors, *Economics Letters* **91** (2006), 21–26.

[11] Linnemann, L. and Schabert, A., *Journal of Money, Credit and Banking* **35** (2003), 911–929.

[12] Mankiw, G.: The savers-spenders theory of fiscal policy, *The American Economic Review*, **90** (2000), 120–125.

[13] McKeirnan, B.: Consumption and the credit market, *Economics Letter* **51** (1996), 83–88.

Application of Cooperative Game Solution Concepts to a Collusive Oligopoly Game

David Bartl[1]

Abstract. An oligopoly is a market where a couple of large producers supply some goods. If the oligopoly is collusive, the producers form coalitions. Then, within each of the coalitions, the producers wish to divide their total profit among themselves. They could use a cooperative transferable utility game solution concept, such as the core, if the game were in the coalitional form. This, however, is not the case. In this paper, we propose an approach to overcome that difficulty: converting the collusive oligopoly into the partition function form, we show how the known cooperative game solution concepts (core, bargaining set) can be applied to that game. Actually, the proposed approach is suitable not only for a collusive oligopoly, but, under some assumptions, for any cooperative strategic form game.

Keywords: cooperative games, partition function form games, solution concepts, core, bargaining set.

JEL classification: C71, L13
AMS classification: 91A12

1 Introduction

Let us consider a game in the *strategic* (or *normal*) *form*: let $N = \{1, \ldots, n\}$ be the set of the *players*, let $X_1, \ldots, X_n$ be the *strategy spaces* of the players, and let $F_1, \ldots, F_n$ be their *payoff functions*. Each X_j is a non-empty set and each F_j is a real function defined on the Cartesian product $X_1 \times \cdots \times X_n$.

For a player $j \in N$, the set X_j consists of all the decisions (called *strategies*) which the player j can make. At a moment, each of the players $j \in N$ picks up a strategy $x_j \in X_j$ and receives the amount of $F_j(x_1, \ldots, x_n)$ of some utility (such as money).

An example of a game in the strategic form is an *oligopoly*. It is a market where a couple of large producers supply some goods. Each of the producers has enough power to influence the market by its decisions. We shall describe the *Cournot model* [3] of an oligopoly here. Let $N = \{1, \ldots, n\}$ be the set of the oligopolists. They supply one kind of some goods, product or commodity (such as metal, grain, oil, etc.). For $j \in N$, let $L_j > 0$ be the *production limit* of the oligopolist j, i.e. the maximum amount of the goods the oligopolist is able to supply to the market. Then $X_j = \langle 0, L_j \rangle$, a closed interval, is the oligopolist's strategy space. Now, each of the oligopolists decides to supply some amount $x_j \in X_j$ of the goods to the market. Hence, the total supply of the goods is $s = \sum_{j=1}^{n} x_j$. Then an internal mechanism of the market, which effects so that the market clears (the supply equals the demand for the goods), establishes the price $p(s)$ per a unit of the goods. The function p is the *price* (or *inverse demand*) *function* of the oligopoly. The price function p is a real function defined on the closed interval $\langle 0, L \rangle$ where $L = \sum_{j=1}^{n} L_j$ is the maximum total supply of the goods to the market. Then $x_j p(s)$, the supply multiplied by the unit price, is the revenue of the oligopolist j. However, the oligopolist faces some production costs connected with the supply of the amount x_j of the goods. The oligopolist's production costs are $c_j(x_j)$ where c_j is the *cost function* of the oligopolist j. The cost function c_j is a real function defined on the interval $X_j = \langle 0, L_j \rangle$. Finally, the oligopolist's net profit is $F_j(x_1, \ldots, x_n) = x_j p(s) - c_j(x_j)$ for $j \in N$ where $s = \sum_{j=1}^{n} x_j$.

Given a strategic form game (oligopoly), we say that $[x_1^*, \ldots, x_n^*] \in X_1 \times \cdots \times X_n$ is a point of the *Nash equilibrium* iff, for each $j \in N$, it holds $F_j(x_1^*, \ldots, x_{j-1}^*, x_j, x_{j+1}^*, \ldots, x_n^*) \leq F_j(x_1^*, \ldots, x_{j-1}^*,$

[1]University of Ostrava, Faculty of Science, Department of Mathematics, 30. dubna 22, 701 03 Ostrava, Czech Republic, e-mail: bartl@osu.cz

$x_j^*, x_{j+1}^*, \ldots, x_n^*$) for all $x_j \in X_j$. The Nash equilibrium is also termed the *Cournot equilibrium* or the *Cournot-Nash equilibrium* if the strategic form game under consideration is a Cournot oligopoly.

Let us assume the given strategic form game (oligopoly) is non-cooperative in the following sense: (A) each player j chooses the decision $x_j \in X_j$ independently of the other players and, at the same time, maximizes own profit $F_j(x_1, \ldots, x_n)$, disregarding the payoffs of the remaining players; (B) if the situation has "settled down", i.e., is in a Nash equilibrium, then the players do not react at all (or react very slowly) if only one of them tries to change the equilibrium state; as the player's payoff F_j does not increase, the player who tried to change the situation is returned back to the equilibrium state. Provided these two assumptions (A) and (B) are satisfied and the game has exactly one point of the Nash equilibrium, let us assume that the situation arrives there. Then, if we manage to find the unique point $[x_1^*, \ldots, x_n^*]$ of the Nash equilibrium, we shall also know the individual payoffs $F_j(x_1^*, \ldots, x_n^*)$, which the players will receive.

The question whether a point of the Nash equilibrium exists is addressed by the original papers of Nash [11, 12], by the classical paper of Nikaidô and Isoda [13], or by more recent papers, e.g. [14]. Conditions for the existence and uniqueness of the Cournot equilibrium can be found, e.g., in the papers [8, 5]. (See also [4].) Especially interesting are the papers of Szidarovszky and Yakowitz [16, 17] and the book [15].

Now, let us consider that the given strategic form game is cooperative (the given oligopoly is collusive). That is, the players form coalitions. A *coalition* is any subset of the set of the players N. Let $S = \{j_1, \ldots, j_{n_S}\} \subseteq N$ be a coalition which has emerged. The members $j_1, \ldots, j_{n_S}$ of the coalition coordinate their decisions in order to maximize their total payoff. Naturally, the coalition's strategy space X_S comprises the Cartesian product $X_{j_1} \times \cdots \times X_{j_{n_S}}$; when the players choose a strategy from the Cartesian product, the total payoff F_S of the coalition is the sum $\sum_{j \in S} F_j$ of the individual payoffs of its members. However, the strategy space X_S may contain yet additional strategies, which are not in the Cartesian product, i.e., they are feasible only if the players join because a single player or a smaller group of the players may not have enough power to realise those decisions.

The situation simplifies if the considered cooperative game is a collusive Cournot oligopoly. Given a coalition $S = \{j_1, \ldots, j_{n_S}\} \subseteq N$, the oligopolists $j_1, \ldots, j_{n_S}$ simply join their production capacities so that the coalition's strategy space X_S reduces to the closed interval $\langle 0, L_S \rangle$ where $L_S = \sum_{j \in S} L_j$. When the coalition decides to supply some amount $x_S \in X_S$ of the goods to the market, the oligopolists $j_1, \ldots, j_{n_S}$ will allocate their production to their most efficient plants. Hence, the coalition's cost function c_S is calculated as $c_S(x_S) = \min\{ \sum_{j \in S} c_j(x_j) : \sum_{j \in S} x_j = x_S \text{ with } x_j \in X_j \text{ for } j \in S \}$, provided that the minimum exists. Therefore, the coalition's net profit is $F_S = x_S p(s) - c_S(x_S)$ where s is the total supply of the goods to the market.

It is definitely beyond the scope of this paper to study the process of the formation of coalitions. Following [10], we shall simply assume that a coalition structure will "crystallize". A *coalition structure* is any partition of the set N. In other words, a coalition structure is a collection $\mathcal{S} = \{S_1, \ldots, S_\nu\}$ of coalitions such that $\bigcup_{\iota=1}^{\nu} S_\iota = N$ and $S_{\iota'} \cap S_{\iota''} = \emptyset$ iff $\iota' \neq \iota''$ for all $\iota', \iota'' = 1, \ldots, \nu$.

Thus, we obtain another strategic form game where the set of the players is the set of the established coalitions $\mathcal{S} = \{S_1, \ldots, S_\nu\}$, their strategy spaces are $X_{S_1}, \ldots, X_{S_\nu}$ and their payoff functions are $F_{S_1}, \ldots, F_{S_\nu}$. It is quite natural to assume that, in this game, the coalitions will behave in a mutually non-cooperative way, making the above assumptions (A) and (B) hold. Hence, if there exists exactly one point $[x_{S_1}^*, \ldots, x_{S_\nu}^*]$ of the Nash equilibrium in the game, the total payoffs $F_{S_\iota}(x_{S_1}^*, \ldots, x_{S_\nu}^*)$ of the coalitions can be determined for $\iota = 1, \ldots, \nu$.

Now, let the members $j_{\iota,1}, \ldots, j_{\iota,n_\iota}$ of an established coalition $S_\iota = \{j_{\iota,1}, \ldots, j_{\iota,n_\iota}\} \in \mathcal{S}$ wish to divide their total profit F_{S_ι} among themselves. If the game were in the coalitional form (see Section 3), then the members could use, e.g., the concept of the core [7], the bargaining set [10], or another transferable utility (TU) game solution concept. That, however, is not the case: the cooperative game under consideration is in the strategic form.

In this paper, in order to apply TU-game solution concepts to the given cooperative strategic form game, we propose the following approach: First, to convert the cooperative strategic form game into the partition function form (Section 2). And, then, to apply cooperative TU-game solution concepts to the partition function form game (Section 3). The approach proposed here extends and generalises the author's earlier idea which originally appeared in [1].

2 Conversion of a cooperative strategic form game into a partition function form game

Let $N = \{1, \dots, n\}$ be the set of the players. Let $\mathfrak{S}$ be the set of all coalition structures $\mathcal{S}$ that the players can form. Recall that a *partition function form game*, in the famous sense of Thrall and Lucas [18], is given by a *partition function* $\mathcal{F}$ which is defined on the set of all coalition structures $\mathfrak{S}$ and, to each coalition structure $\mathcal{S} \in \mathfrak{S}$, it assigns a function $\mathcal{F}_{\mathcal{S}} \colon \mathcal{S} \to \mathbb{R}$. For a coalition $S \in \mathcal{S}$, the value $\mathcal{F}_{\mathcal{S}}(S)$ is the total payoff that the established coalition S will receive.

Given a cooperative game in the strategic form, it is easy to convert it into the partition function form: Let $\mathcal{S} = \{S_1, \dots, S_\nu\} \in \mathfrak{S}$ be a coalition structure, which the players have formed. Let $X_{S_1}, \dots, X_{S_\nu}$ and $F_{S_1}, \dots, F_{S_\nu}$ be the strategy spaces and the payoff functions, respectively, of the coalitions. Let us assume that, *for any coalition structure* $\mathcal{S} \in \mathfrak{S}$*, there exists exactly one Nash equilibrium in the game among the coalitions.* Let $[x^*_{S_1}, \dots, x^*_{S_\nu}] \in X_{S_1} \times \cdots \times X_{S_\nu}$ be the unique point of the Nash equilibrium. We put $\mathcal{F}_{\mathcal{S}}(S_\iota) = F_{S_\iota}(x^*_{S_1}, \dots, x^*_{S_\nu})$ for $\iota = 1, \dots, \nu$. The conversion has been described thus.

The assumption of the existence and uniqueness of the Nash equilibrium for any coalition structure $\mathcal{S} \in \mathfrak{S}$ can be seen quite restrictive. However, it can be met in the case of a Cournot oligopoly for example. It is not difficult to show [16, 17, 15] that if the price function p of the oligopoly is linear and decreasing (so $p(s) = as + b$ for some $a < 0$ and $b > 0$) and the oligopolists' cost functions c_j are non-increasing and convex, then the oligopoly possesses exactly one Cournot equilibrium. Moreover, it is an exercise to show that if the cost functions c_j are convex, then, for any coalition $S \subseteq N$, the coalition's cost function $c_S(x_S) = \min\{ \sum_{j \in S} c_j(x_j) : \sum_{j \in S} x_j = x_S \text{ with } x_j \in X_j \text{ for } j \in S \}$ is also convex. Hence, there exists a unique Cournot equilibrium for any coalition structure $\mathcal{S} \in \mathfrak{S}$ in the non-cooperative oligopolistic game among the coalitions.

3 Application of Cooperative TU-Game Solution Concepts to a Partition Function Form Game

Let us consider a cooperative game with transferable utility (TU) in the *coalitional form*: Let $N = \{1, \dots, n\}$ be the set of the players. A *coalition* is any subset of the set of the players. Then $\mathcal{P}(N) = \{ S : S \subseteq N \}$, the potency set of the set N, is the collection of all coalitions which can be formed. Finally, let $v \colon \mathcal{P}(N) \to \mathbb{R}$ with $v(\emptyset) = 0$ be the *coalitional* (or *characteristic*) *function* of the game.

When a coalition $S \subseteq N$ is formed, it receives the amount of $v(S)$ units of some utility. It is assumed here that the utility is *transferable*, that is, the members of the coalition S can divide the amount among themselves.

Let us consider that the players have formed a coalition structure $\mathcal{S} = \{S_1, \dots, S_\nu\}$. Then $v(S_\iota)$ is the payoff that the coalition S_ι receives for $\iota = 1, \dots, \nu$. Now, the question, which the cooperative game theory studies, is how will the members of the coalitions S_ι divide their payoffs $v(S_\iota)$ among themselves.

The division of the profit among the players is described by the payoff vector. A *payoff vector* is any n-tuple $a = [a_1, \dots, a_n] \in \mathbb{R}^n$. The number a_j stands for the amount which is allotted to the player $j \in N$.

Several solution concepts – such as the core [7] or the bargaining set [10] – were proposed to address the question. Recall that a *solution concept* is a mapping that, to a given coalitional function $v \colon \mathcal{P}(N) \to \mathbb{R}$ with $v(\emptyset) = 0$ and a given coalition structure $\mathcal{S}$, assigns a set of payoff vectors; sometimes, it assigns a collection of sets of payoff vectors or just a single payoff vector (in the case of the von Neumann-Morgenstern solution or the Shapley value, respectively; we shall not deal with these solution concepts in this paper).

Here, given a coalition structure $\mathcal{S}$, we would like to apply those solution concepts to a partition function form game $\mathcal{F}$. In the following, we shall recall and contemplate the solution concept of the core and that of the bargaining set.

Given a coalitional function $v \colon \mathcal{P}(N) \to \mathbb{R}$ with $v(\emptyset) = 0$ and a coalition structure $\mathcal{S} = \{S_1, \dots, S_\nu\}$, the *core* of the game is the set $\mathcal{C} = \{ a \in \mathbb{R}^n : \sum_{j \in S_\iota} a_j = v(S_\iota) \text{ for } S_\iota \in \mathcal{S} \text{ and } \sum_{j \in S} a_j \geq v(S) \text{ for all } S \in \mathcal{P}(N) \setminus \mathcal{S} \}$.

Now, given the coalition structure $\mathcal{S} = \{S_1, \ldots, S_\nu\}$, as above, and recalling the motivation stated in the Introduction, let us apply the concept of the core to a partition function form game $\mathcal{F}$. We can indeed formulate the equalities that are a part of the description of the core: let $\sum_{j \in S_\iota} a_j = \mathcal{F}_{\mathcal{S}}(S_\iota)$ for $S_\iota \in \mathcal{S}$. The equalities mean that each of the established coalitions S_ι divides all its profit $v(S_\iota)$ among its members. Nonetheless, how about the inequalities $\sum_{j \in S} a_j \geq v(S)$ for $S \in \mathcal{P}(N) \setminus \mathcal{S}$? Do we need the inequalities – what do they mean?

The inequalities $\sum_{j \in S} a_j \geq v(S)$ for $S \in \mathcal{P}(N) \setminus \mathcal{S}$ are the *conditions of group stability*. Let us consider a coalition $S \in \mathcal{P}(N) \setminus \mathcal{S}$. That is, the coalition does actually *not* exist, but *could* potentially be formed. Should the respective inequality not hold, so we would have $\sum_{j \in S} a_j < v(S)$, then the coalition S would have a good reason to form because its total payoff $v(S)$ will be higher that the present total payoff $\sum_{j \in S} a_j$ of its members. That is, if $\sum_{j \in S} a_j < v(S)$, then the present coalition structure $\mathcal{S}$ is instable and the new coalition S will emerge.

Let us continue that thoughts: When the new coalition $S \in \mathcal{P}(N) \setminus \mathcal{S}$ emerges, what happens with the coalition structure $\mathcal{S}$? We assume that a new coalition structure $\mathcal{S}_S$, containing S, will form shortly after the emergence of the coalition S. Which particular coalition structure $\mathcal{S}_S \in \mathfrak{S}$ will form, i.e., which coalitions it will contain, depends on the chosen approach. In this paper, we mention the γ-approach and the δ-approach of Hart and Kurz [6].

If we assume the γ-approach, then the new coalition structure will be $\mathcal{S}_S = \{S\} \cup \{\, S_\iota \in \mathcal{S} : S_\iota \cap S = \emptyset \,\} \cup \{\, \{j\} : \exists S_\iota \in \mathcal{S} \colon j \in S_\iota \setminus S \,\}$. In words, the new coalition structure $\mathcal{S}_S$ contains the new coalition S, all the formerly established coalitions $S_\iota \in \mathcal{S}$ not affected by the departure ($S_\iota \cap S = \emptyset$), but the other coalitions ($S_\iota \cap S \neq \emptyset$) split into singletons $\{j\}$.

If we assume the δ-approach, then the new coalition structure will be $\mathcal{S}_S = \{S\} \cup \{\, S_\iota \setminus S : S \not\supseteq S_\iota \in \mathcal{S} \,\}$. In words, the new coalition structure $\mathcal{S}_S$ contains the new coalition S, all the formerly established coalitions $S_\iota \in \mathcal{S}$ not affected by the departure ($S_\iota \cap S = \emptyset$), but the remaining non-empty parts $S_\iota \setminus S$ of the other coalitions ($S_\iota \cap S \neq \emptyset$) stay intact.

Now, it is easy to formulate the conditions of group stability for the partition function form game $\mathcal{F}$ and the established coalition structure $\mathcal{S}$. We write $\sum_{j \in S} a_j \geq \mathcal{F}_{\mathcal{S}_S}(S)$ for $S \in \mathcal{P}(N) \setminus \mathcal{S}$. To conclude, we define the *core* of the partition function form game $\mathcal{F}$ with respect to the coalition structure $\mathcal{S} = \{S_1, \ldots, S_\nu\}$ to be the set $\mathcal{C} = \big\{\, a \in \mathbb{R}^n : \sum_{j \in S_\iota} a_j = \mathcal{F}_{\mathcal{S}}(S_\iota)$ for $S_\iota \in \mathcal{S}$ and $\sum_{j \in S} a_j \geq \mathcal{F}_{\mathcal{S}_S}(S)$ for all $S \in \mathcal{P}(N) \setminus \mathcal{S} \,\big\}$.

Note that, in the definition of the core, the payoffs $\mathcal{F}_{\mathcal{S}_S}(S')$ of the other coalitions $S' \in \mathcal{S}_S \setminus \{S\}$ from the new coalition structure $\mathcal{S}_S$ are immaterial to us. However, when a coalition $S \in \mathcal{P}(N) \setminus \mathcal{S}$ departs, neither the original definition of the core of a coalitional form game considers what happens with the payoffs of the other coalitions.

We shall deal with the concept of the bargaining set in the rest of this section. We shall recall the concept of the imputation, objection, and counterobjection first.

Let a coalitional function $v \colon \mathcal{P}(N) \to \mathbb{R}$ with $v(\emptyset) = 0$ and a coalition structure $\mathcal{S} = \{S_1, \ldots, S_\nu\}$ be given. Then the set of the *imputations* of the game is the set $X = \big\{\, a \in \mathbb{R}^n : \sum_{j \in S_\iota} a_j = v(S_\iota)$ for $S_\iota \in \mathcal{S}$ and $a_j \geq v(\{j\})$ for all $j \in N \,\big\}$. As above, the equalities $\sum_{j \in S_\iota} a_j = v(S_\iota)$ mean that each of the established coalitions S_ι divides all its profit $v(S_\iota)$ among its members. The inequalities $a_j \geq v(\{j\})$ are the *conditions of individual rationality*. For a $j \in N$, the one-player coalition $\{j\}$ does actually not exist (unless $\{j\} \in \mathcal{S}$), but, if $a_j < v(\{j\})$, i.e., the player j receives less than the player can obtain by forming own independent coalition, then the present coalition structure $\mathcal{S}$ is instable and the coalition $\{j\}$ will emerge.

Consider two distinct players $k, l \in S_\iota \in \mathcal{S}$, $k \neq l$, from an established coalition S_ι. Let $a \in X$ be an imputation under the consideration.

An *objection* of the player k against l at the imputation a is a pair (K, b) where $K \subseteq N$ is a coalition such that $k \in K \not\ni l$ and $b \in \mathbb{R}^K$ is such that $\sum_{j \in K} b_j = v(K)$ and $b_j > a_j$ for all $j \in K$. That is, the coalition K does actually not exist, but has the potential to form because all its new members will receive higher payoffs than under the current division a of the profit. Note that the concept of the objection does not concern with the payoffs of the players outside the coalition K if the coalition separates.

A *counterobjection* of the player l to the objection (K, b) of k against l at a is a pair (L, c) where $L \subseteq N$ is a coalition such that $l \in L \not\ni k$ and $l \in \mathbb{R}^L$ satisfies $\sum_{j \in L} c_j = v(L)$ with $c_j \geq b_j$ for all $j \in L \cap K$ and $c_j \geq a_j$ for all $j \in L \setminus K$. So, it is assumed now that the coalition K has really emerged.

The coalition L does not actually exist, but again has the potential to form because the members from K as well as the new members from outside K will receive the same or higher payoffs than under the current division b or a, respectively, of the profit. Note that neither the concept of the counterobjection concerns with the payoffs of the players outside the coalition L if it forms.

We say that an objection is *justified* iff there is no counterobjection to it. Finally, the *bargaining set* is the set of all imputations $a \in X$ such that there does not exist any justified objection at a. That is, the bargaining set is the set $\mathcal{M}_1^i = \{\, a \in X : \forall S_\iota \in \mathcal{S}\ \forall k, l \in S_\iota,\ k \neq l,\ \forall (K, b),\ (K, b)$ is an objection of k against l at a, $\exists (L, c)$, (L, c) is a counterobjection of l to (K, b) of k against l at $a \,\}$.

Now, having understood the concept of the core earlier, it is easy to restate the above definitions in the setting of a partition function form game $\mathcal{F}$. Let $\mathcal{S} = \{S_1, \ldots, S_\nu\}$ be the established coalition structure. Recall that $\mathcal{S}_S$ denotes the coalition that will form if the coalition $S \in \mathcal{P}(N) \setminus \mathcal{S}$ decides to depart. We can adopt several approaches (such as the γ-approach or the δ-approach) to define $\mathcal{S}_S$. (We put $\mathcal{S}_S = \mathcal{S}$ if $S \in \mathcal{S}$.)

We define the set of the *imputations* to be the set $X = \big\{\, a \in \mathbb{R}^n : \sum_{j \in S_\iota} a_j = \mathcal{F}_{\mathcal{S}}(S_\iota)$ for $S_\iota \in \mathcal{S}$ and $a_j \geq \mathcal{F}_{\mathcal{S}_{\{j\}}}(\{j\})$ for all $j \in N \,\big\}$.

Let $k, l \in S_\iota \in \mathcal{S}$, $k \neq l$, be two distinct players from an established coalition and let $a \in X$ be an imputation. We define an *objection* of the player k against l at the imputation a to be a pair (K, b) where $K \subseteq N$ is such that $k \in K \not\ni l$ and $b \in \mathbb{R}^K$ is such that $\sum_{j \in K} b_j = \mathcal{F}_{\mathcal{S}_K}(K)$ and $b_j > a_j$ for all $j \in K$. And we define a *counterobjection* of the player l to the objection (K, b) of k against l at a to be a pair (L, c) where $L \subseteq N$ is such that $l \in L \not\ni k$ and $c \in \mathbb{R}^L$ satisfies $\sum_{j \in L} c_j = \mathcal{F}_{(\mathcal{S}_K)_L}(L)$ with $c_j \geq b_j$ for all $j \in L \cap K$ and $c_j \geq a_j$ for all $j \in L \setminus K$.

Finally, we define the *bargaining set* to be the set of all imputations $a \in X$ such that there does not exist any justified objection at a, i.e., to be the set $\mathcal{M}_1^i = \{\, a \in X : \forall S_\iota \in \mathcal{S}\ \forall k, l \in S_\iota,\ k \neq l,\ \forall (K, b)$, (K, b) is an objection of k against l at a, $\exists (L, c)$, (L, c) is a counterobjection of l to (K, b) of k against l at $a \,\}$.

4 Conclusions

We considered a cooperative game in the strategic form. The classical solution concepts (the core, the bargaining set, etc.), being defined for coalitional form games, cannot be applied to that game directly. Therefore, under the assumption of the existence and uniqueness of the Nash equilibrium, we proposed in Section 2 to convert the cooperative strategic form into a partition function form game.

We noted in Section 2 that if the price function of a Cournot oligopoly is linear and decreasing and the cost functions of the oligopolists are convex and non-increasing, then there exists exactly one Cournot equilibrium, whence the proposed conversion is possible. (See also [9].) It is a motivation of further research to find more general conditions under which there exists (exactly one) Cournot equilibrium in the oligopoly.

In Section 3, we showed how to apply the concept of the core and that of the bargaining set, which are defined for coalitional form games, to a partition function form game. Consequently, they can be applied to the original cooperative strategic form game (under the assumption of the existence and uniqueness of the Nash equilibrium), such as the collusive Cournot oligopoly.

For the lack of the space, we did not deal with other popular solution concepts (the von Neumann-Morgenstern solution, the kernel, the nucleolus, or the Shapley value) in Section 3. They could be applied analogously.

Note that if the considered partition function form game $\mathcal{F}$ is the result of the conversion of a cooperative strategic form game, if $\mathcal{S} = \{N\}$, i.e., the coalition structure contains only the grand coalition of the players, and if we assume the γ-approach, then our definition of the core of the partition function form game yields precisely the concept of the γ-core of Chander and Tulkens [2]. Nonetheless, our approach is more general in the sense that we define the core for *any* coalition structure $\mathcal{S}$.

Actually, it was essential for the conversion described in Section 3 to decide upon the approach which coalition structure $\mathcal{S}_S$ will form when a coalition $S \in \mathcal{P}(N) \setminus \mathcal{S}$ departs from $\mathcal{S}$. The application of the γ-approach, the δ-approach, etc., results in the concept of the γ-core, the δ-core, etc., the concept of the

γ-bargaining set, the δ-bargaining set, etc., etc. While our concept of the γ-core is more general than that of [2], as already mentioned, the concept of the γ-bargaining set or the δ-bargaining set is, according to the author's best knowledge, new.

References

[1] Bartl, D. Translation of a Collusive Oligopoly Game to a Coalitional Form Game. In: *Proceedings of the 19th International Conference Mathematical Methods in Economics 2001*, Vysoká škola ekonomická v Praze, 2001, 1–6. ISBN 80-245-0196-1.

[2] Chander, P., and Tulkens, H.: The Core of an Economy with Multilateral Environmental Externalities, *International Journal of Game Theory* **26** (1997), 379–401.

[3] Cournot, A. A.: *Recherches sur les principes mathématiques de la théorie des richesses*. L. Hachette, Paris, 1838. English translation by N. T. Bacon: *Researches into the Mathematical Principles of the Theory of Wealth*. Macmillan, New York, 1897, 1927. Reprinted: Augustus M. Kelly, New York, 1960, 1964, 1971. ISBN 0-678-00066-2.

[4] Everhart, C.: Cournot Oligopoly and Concavo-Concave Demand, *University of Zurich. Department of Economics: Working Paper Series*, Working Paper No. **16** (April 2011).

[5] Gaudet, G., and Salant, S. W.: Uniqueness of Cournot Equilibrium: New Results from Old Methods, *Review of Economic Studies* **58** (1991), 399–404.

[6] Hart, S., and Kurz, M.: Endogenous Formation of Coalitions, *Econometrica* **51** (1983), 1047–1064.

[7] Kannai, Y.: The Core and Balancedness. In: *Handbook of Game Theory with Economic Applications. Volume 1* (R. J. Aumann, and S. Hart, eds.), North-Holland, Amsterdam, 1992, 355–395. ISBN 0-444-88098-4.

[8] Kolstad, C. D., and Mathiesen, L.: Necessary and Sufficient Conditions for Uniqueness of a Cournot Equilibrium, *Review of Economic Studies* **LIV** (1987), 681–690.

[9] Lardon, A.: The γ-core in Cournot oligopoly TU-games with capacity constraints, *Theory and Decision* **72** (2011), 387–411.

[10] Maschler, M.: The Bargaining Set, Kernel, and Nucleolus. In: *Handbook of Game Theory with Economic Applications. Volume 1* (R. J. Aumann, and S. Hart, eds.), North-Holland, Amsterdam, 1992, 591–667. ISBN 0-444-88098-4.

[11] Nash, J., Jr.: Equilibrium Points in N-Person Games, *Proceedings of the National Academy of Sciences of the United States of America* **36** (1950), 48–49.

[12] Nash, J.: Non-Cooperative Games, *Annals of Mathematics* **54** (1951), 286–295.

[13] Nikaidô, H., and Isoda, K.: Note on Noncooperative Convex Games, *Pacific Journal of Mathematics* **5** (1955), 807–815.

[14] Nishimura, K., and Friedman, J.: Existence of Nash Equilibrium in n Person Games without Quasi-Concavity, *International Economic Review* **22** (1981), 637–648.

[15] Okuguchi, K., and Szidarovszky, F.: *The Theory of Oligopoly with Multi-Product Firms*. Springer-Verlag, Berlin, 1990. ISBN 3-540-52567-X.

[16] Szidarovszky, F., and Yakowitz, S.: A New Proof of the Existence and Uniqueness of the Cournot Equilibrium, *International Economic Review* **18** (1977), 787–789.

[17] Szidarovszky, F., and Yakowitz, S.: Contributions to Cournot Oligopoly Theory, *Journal of Economic Theory* **28** (1982), 51–70.

[18] Thrall, R. M., and Lucas, W. F.: n-Person Games in Partition Function Form, *Naval Research Logistics Quarterly* **10** (1963), 281–298.

Use of the three-point PERT estimate in Critical Chain method

Jan Bartoška[1], Tomáš Šubrt[2]

Abstract. The paper proposes the stochastic modification of the Critical Chain method. The authors design new computation procedure for time buffers with the use of the three-point PERT estimate. The proposed procedure in Critical Chain method reflects the human factor in scheduled tasks. The original Goldratt time estimate purging works with 50% purging for all tasks in the project. However the tasks in the project are usually of various kinds with different dependency on human factor. There should be a relevant value of concealed time reserve for each concrete task. Within real projects the three-point PERT estimate is often adapted to suit specific requirements and project needs. Possible modification of the three-point PERT estimate offers many options which incorporate the human factor into the projects. Usually, only the pessimistic and the optimistic parameters of the three-point PERT estimate are precisely set. The pessimistic and optimistic parameters can be used for the estimate of a concealed time reserve. The concealed time reserve is needed for the time estimate purging within the Critical Chain method. The correct time estimate purging with various concealed time reserve for each task can bring saving costs of the project. The handling of human factor is still an important issue in project management.

Keywords: project management; Critical Chain method; thee-point PERT Estimate; concealed time reserve; time estimate purging; work effort; Student Syndrome.

JEL Classification: C61.
AMS Classification: 90B99.

1 Introduction

Even after the first decade of the twenty-first century it is evident that still many projects exceed their deadlines and their budgets. The impact of human factor on every project and its objectives is indisputable. And it is the human factor that is a frequent denominator of evident or hidden causes of project failure. Applying international standards (such as PMI, IPMA, and Prince2) in project management restricts human factor influence to some degree or in a certain phase of the project only. Not even thorough application of the latest mathematical methods and approaches leads to an essential elimination of the human factor impact. The methods or approaches are still unable to capture the human factor impact. The methods which can enable work with the human factor impact are the PERT method and Critical Chain method. In specialist literature, both methods are constantly developed and modified, particularly in [1], [6], [8], [9], [12], [13], [14] or [15].

Deficiencies and possible modifications of a commonly used approach for the computation of the mean value and time estimate dispersion in PERT method is discussed for instance in [8] and [13]. The author of [8] however, deals with the use of rectangular beta distribution and its possible advantages for application. Another approach to the modification of a current way of computation using PERT method is offered by [13] who proposes his own way of mean value approximation and time estimate dispersion and compares them with the existing ways. Both articles work with mathematical apparatus only and propose the modification of either beta distribution or the mean value and dispersion computation. In their results, the above-mentioned articles do not directly offer the impact of the human factor in determining the duration time estimate. The way of application and possible modification of the PERT method is further dealt with in for example [1], [6] or [12]. A key element in the PERT method is its three-point estimate which relies on a presupposed beta distribution of individual activity duration time. In practice there exist a number of application alternatives for three-point

[1] Ing. Jan Bartoška, Ph.D., bartoska@pef.czu.cz, Department of Systems Engineering,
Faculty of Economics and Management, Czech University of Life Sciences Prague, Kamýcká 129, 165 21 Prague 6, Czech Republic.
[2] Doc. Ing. Tomáš Šubrt, Ph.D., subrt@pef.czu.cz, Department of Systems Engineering,
Faculty of Economics and Management, Czech University of Life Sciences Prague, Kamýcká 129, 165 21 Prague 6, Czech Republic

estimate without a more profound theoretical framework although the use of the three-point estimate presupposing the beta distribution even in other project management methods and approaches is at hand. The benefit of the three-point estimate consists in the incorporation of the human factor impact into activity duration. The estimate of optimistic and pessimistic duration presents and always will present particular and unequivocal information about the difficulty of the task and presupposed effort of an allocated resource. There are many articles either directly or indirectly dealing with the human factor impact on the project realization; however the articles do not look for solutions or do not incorporate them in their proposals. In projects the human factor impact can be seen as "Student Syndrome" phenomenon. The "Student Syndrome" is discussed for example in [5] and [10]. The phenomenon was one of the basic conditions for Critical Chain method in [7] or also in [5]. The deficiencies of the Critical Chain method are particularly discussed in Raz et al. [14] who stress its contribution and encourage its further development. However, the combination of the three-point PERT estimate and Critical Chain method is discussed only by few.

The following text of the article focuses on the modification of a concealed time reserve and its use for modified computation of time buffers in the Critical Chain method. The article aims to propose the way of three-point estimate use in the Critical Chain method in order to eliminate the impact of the human factor.

2 Material and methods

2.1 Three-point estimate modification

In article [2], authors Bartoška and Šubrt deal with the modification of the three-point PER estimate. A proposed modification proceeds from the human factor influence in project management and lies in shifting a duration distribution peak towards a pessimistic or optimistic estimate. The precondition of the proposal is a variable duration distribution peak for variable work effort. Another approach to the PERT method modification is discussed for example by Hanh [8] or Premachandra [13]. In article [2] the shifting of the duration distribution peak is enabled by a proposed mean value computation. The proposal is based on the elimination of a mode value, i.e. an estimate value of the most frequent activity duration. The shifting direction is determined by a proposed γ parameter with the range of values $(0; +\infty)$. The parameter expresses the impact of the human factor, i.e. "Student Syndrome" on activity duration distribution. Bartoška and Šubrt [2] mention that the more the resource succumbs to the "Student Syndrome", the more an activity duration distribution peak shifts towards a pessimistic estimate. The resource's succumbing to the "Student Syndrome" does not change in a short span of time. The pertinence of the mode estimate, i.e. the position of a distribution peak, will depend on the level of resource's self-criticism which is not taken into account in the PERT method. If the resource knowingly or unknowingly sets the modal parameter in a way different from reality or long-term experience, there will be an imprecise mean value estimate of the duration, which can jeopardize the whole project. To a certain degree, the pertinence of the most frequent activity duration estimate also determines the delay of the activity as well as the delay of the project. Verifying the plausibility and correctness of the duration mode estimate can be an uneasy and impossible task for a project manager. If the project manager has enough information and experience with the resource, s/he can make a better estimation of an activity duration distribution peak than the resource. Unlike the resource, the manager does not succumb to the subjectivity of a resource allocated to the activity. Article [2] therefore proposes the modification of the three-point estimate of activity duration not using a modal value:

$$\mu_{ij} = \frac{\gamma_{ij} b_{ij} + a_{ij}}{1 + \gamma_{ij}} \tag{1}$$

where parameters b_{ij} and a_{ij} in formula (1) are pessimistic and optimistic estimates of activity duration, and parameter γ_{ij} is the level of the "Student Syndrome" influence on an allocated human resource in the activity. The project manager can set the value of parameter γ_{ij} in the range of $(0; +\infty)$ and determine an activity duration distribution peak. A very low value of parameter γ presents the state where the "Student Syndrome" has nearly no influence and the mean value of activity duration approaches an optimistic estimate. An extremely high value of parameter γ expresses absolute influence of the syndrome where the mean value of the activity duration approaches a pessimistic estimate. The more the value of parameter γ approaches 0, the more the resource is knowledgeable in his/her work effort and vice versa. The parameter γ, proposed in article [2], can be further interpreted as an estimate coefficient for maximum load of the resource during activity realization.

2.2 Individual time estimate purging

In his work [3], Bartoška deals with the modification of the present Critical Chain method modification using a wider interpretation of the PERT method beta distribution parameters. The Critical Chain method, discussed particularly by Goldratt [7] or also Leach [10], requires time purging for individual activities. Bartoška [3] proposes the purging of particular activities individually according to their character and the character of their resources. The determination of the three-point estimate parameters for each activity in the project takes place by way of questions and answers between the project manager and a responsible resource. When estimating the parameters, a worker responsible for the activity, i.e. a resource or resource team, succumbs to the "Student Syndrome". This phenomenon, expected in the resource's behaviour, will have influence on the estimate of the most frequent duration (m_{ij}) and the values of pessimistic duration (b_{ij}) of the activity. While for m_{ij} value estimate the resource proceeds especially from his or her experience and historical data and facts, for b_{ij} value estimate the resource succumbs more to his or her expectations. This can be explained by the fear of possible failure and effort to gain time reserve for the activity the resource is responsible for. When estimating m_{ij} value, there exists "Student Syndrome" impact from the viewpoint of the experience obtained by the investigator during his or her practical training only. It can be referred to as concealed influence of the phenomenon. When estimating b_{ij} value, there exist "Student Syndrome" influence from the expected behaviour of a future investigator only, i.e. it can be referred to as a direct influence of the phenomenon. If the human factor influence, i.e. "Student Syndrome", is incorporated in reducing former given activity duration, based on the three-point estimate of arbitrary project activity, a purging value (p_{ij}) can be determined as follows:

$$p_{ij} = \frac{b_{ij} - m_{ij}}{b_{ij}} \tag{2}$$

The derived expression (2) in [3] presents the share of a concealed time reserve of the investigator in a total activity duration *ij*. The computed value (p_{ij}) can be used for activity duration purging (t_{ij}) as follows [3]:

$$t'_{ij} = \left(1 - p_{ij}\right) t_{ij} \tag{3}$$

The precondition for direct shortening of the former activity duration by a particular value is the occurrence of a concealed time reserve of the investigator, which can be expressed as difference between pessimistic time estimate (b_{ij}) and the most frequent time estimate, or modem (m_{ij}). The concealed time reserve of the investigator in relation to the pessimistic time estimate (b_{ij}) expresses time overlap which, in most cases, is not duly used by the resource. The concealed time reserve can be expressed as follows:

$$r_{ij} = b_{ij} - m_{ij} \tag{4}$$

The investigator increases the time estimate by the concealed time reserve which is not used in the end. In practice, this feature is common and is stressed by a right-sided asymmetry in applied PERT method beta distribution. Using the three-point estimate in the Critical Chain method, feeding buffers and a project buffer can be computed taking into account the variety of individual activities, i.e. a greater concern for the human factor influence on individual activities, and consequently in the project as a whole [3].

3 Results and Discussion

3.1 Individual time estimate purging modification

The three-point estimate modification offers the computation of an activity duration mean value without the use of modal value. The activity duration beta distribution peak in proposal [2] will vary depending on the character of the resource and it will be dependent on the human factor influence. The most frequent activity time estimate (m_{ij}), which is the distribution peak value, is represented in the computation of an activity duration mean value by parameter $\gamma_{ij.}$ Therefore it is convenient to modify the computation of a concealed time reserve as well, because in the modification we give up the mode value in favour of reducing human factor influence. Instead of a mode value it is necessary to select other characteristics of the position which would unequivocally express the midpoint of the aggregate. Suitable characteristics are in this case the mean value of activity duration (1). The mode and mean value can differ and they can both express the midpoint of the aggregate in a different interpretation. Mode (m_{ij}) is the most frequent activity duration value. The mean value represents 50% activity duration. Such precondition is theoretical and not always valid in practical project management and for the application of the following proposal it needs to be verified. Of the above-mentioned interpretations the mean value (μ_{ij}) appears to be a more precise middle parameter. Even in his original recommendation, Goldratt [7]

mentions exactly 50% for activity duration purging. However, due to the asymmetry of beta distribution, the final purging (p_{ij}) will correspond to 50% value in several cases only. The duration will be purged by 50% observation, not 50% value. The impact of the human factor can be expressed as a concealed time reserve as follows (3):

$$r_{ij} = b_{ij} - \mu_{ij} \tag{5}$$

where r_{ij} value is a reserve added by a resource and the value of possible activity duration purging. By substituting (1) for (5) the computation of the concealed time reserve can be derived as follows:

$$r_{ij} = \frac{b_{ij} - a_{ij}}{1 + \gamma_{ij}} \tag{6}$$

Expression (6) proceeds from the modification of the three-point estimate and represents the way of concealed time reserve estimate of the resource in the activity. The value of the concealed reserve (r_{ij}) can be used for the determination of a purging value:

$$p_{ij} = \frac{r_{ij}}{b_{ij}} \tag{7}$$

The share of a concealed time reserve (r_{ij}) and pessimistic duration (b_{ij}) should be presented in per cent. The share can purge the original duration which in case of the three-point estimate will be a computed mean value (μ_{ij}).

3.2 Time buffers by the three-point estimate

[7], [10] or [5] summarise the principle of time buffer creation and their use in the Critical Chain method. Usually, shortening original duration for all activities about 50% is presupposed. The subtracted time is then used for the creation of project time buffers. This Critical Chain method approach takes into account all activities in the project at the same level and does not consider possible differences in individual activities and their resources. The time buffer computation can be carried out differently [3], using a concealed time reserves (r_{ij}) and derived purging parameters (p_{ij}). The computation where the three-point PERT estimate and calculated mean value (μ_{ij}) were used will take into account the impact of the human factor on individual activities of a different type and nature. New computation of non-critical paths time buffers using the Critical Chain method is as follows:

$$FB_Q = \sum p_{ij} \ldots_{ij} ; i, j \in Q \tag{8}$$

where feeding buffer FB_Q of a non-critical path Q is determined by the amount of duration mean value multiples (μ_{ij}) with the level of purging (p_{ij}). The computation incorporates those activities which occur on the non-critical path Q only. The computation of the time buffer of a critical path will be similar:

$$PB_E = \sum p_{ij} \ldots_{ij} ; i, j \in E \tag{9}$$

In (9) the project time buffer on a critical path is denoted as PB_E taking into account the fact that the computation comprises only those project activities lying on the critical path E. For the Critical Chain method the purging of activity duration is carried out identically for the whole project. Placing feeding buffers at the end of the non-critical paths and placing project buffers at the end of the project can be done any time. The time buffers can be placed before or after finding a critical path. This is caused by the fact that all time values for computation in the project time schedule are purged about 50%, which does not affect determining the critical path. Critical activities remain critical even after inserting the time buffers. The only change which takes place during the original Critical Chain method computation is the shortening of all time data to the half of their value. However, these described preconditions cease to be valid for the mentioned proposal of the time buffer computation when using the three-point estimate. It is necessary to consider omitting a critical path where the length of a non-critical path with its feeding buffer could become longer than the length of a corresponding critical path segment. The length of the corresponding critical path segment (E_Q), which starts in the place of the beginning and finishes in the place of the end of a non-critical path Q, expresses maximum time for the realization of a non-critical path even with its feeding buffer. Should the non-critical path be delayed and its feeding buffer, based on (8), exhausted, a corresponding segment of a critical path could be omitted. This is

supported by the fact that the above-mentioned proposal of the time buffer computation works with individual activity purging. The activities on a non-critical path can be purged more than the activities on a given segment of a critical path. Then the feeding buffer could grow up enormously. Therefore another way of computing feeding buffers is necessary:

$$FB_Q = \sum \dots_{rs} (1 - p_{rs}) - \sum \dots_{ij} (1 - p_{ij}); i, j \in Q; r, s \in E_Q \tag{10}$$

where the feeding buffer FB_Q of a non-critical path Q is determined by the difference of purged duration amount on segment E_Q of a critical path E and purged duration amount on a non-critical path Q. In sum with non-critical path length, the size of the feeding buffer will always be at least as big as a corresponding segment of a critical path. The critical path will never be omitted.

4 Conclusion

The article deals with possible use of the three-point PERT estimate in the Critical Chain method. The purpose of such connection is to eliminate the impact of human factor. The article presents the human factor impact as "Student Syndrome". Its elimination can be achieved by the use of a modified three-point estimate where an allocated resource is not determined by an activity duration distribution peak (mode) but by a project manager. The activity duration distribution peak can be expressed in dependence on a "Student Syndrome" impact on the allocated resource, which presupposes the existence of a concealed time reserve implicitly incorporated in the duration time estimate made by the resource. The authors of the paper interpret the concealed time reserve of the resource as the difference between the middle distribution characteristics and pessimistic duration estimate. The paper further presents a proposal for the computation of feeding buffers and a project buffer within the Critical Chain method. The proposal of the time buffer computation is based on the use of the concealed time reserve and pessimistic duration estimate as individual purging for particular project activities. The proposals and preconditions mentioned in the article are theoretical without practical computation or example. In the paper, the authors proceed from their previous work and research done on the subject as well as from previously published specialist literature.

References

[1] Azaron, A. et al.: A multi-objective resource allocation problem in PERT networks. *European Journal of Operational Research* **172** (2006).

[2] Bartoška, J., Šubrt, T.: Modification of the Three-point PERT Estimate for Practical Use. In: *Mathematical Methods in Economics*. University of Economics, Prague, 2011, 23-28.

[3] Bartoška, J.: *Integration of Stochactic Approaches and TOC Based Methods in Project Management*. Czech University of Live Sciences in Prague, 2010. Thesis.

[4] Clark, C.E.: The PERT model for the distribution of an activity time. *Operations Research* **10** (1962), 405–406.

[5] Cook, S. C.: *Applying Critical Chain to Improve the Management of Uncertainty in Projects*. Massachusetts Institute of Technology, 1998. Thesis (Online).

[6] Cruz, S. et al.: Stochastic models alternative to the classical PERT for the treatment of the risk: mesokurtic and of constant variance. *Central European Journal of Operational Research* **7** (1999).

[7] Goldratt, E.M.: *Critical Chain*. North River Press, Great Barrington, MA, 1997.

[8] Hanh, E. D.: Mixture densities for project management activity times: A robust approach to PERT. *European Journal of Operational Research* **188** (2008).

[9] Healy, L. Th.: Activity Subdivision and PERT Probability Statements. *Operations Research* **9** (1961).

[10] Leach, L. P.: Critical Chain Project Management Improves Project Performance. *Project Management Journal* **30** (1999), 39-51.

[11] Malcolm, D. G., Roseboom, J.H., Clark, C.E., Fazar, W.: Application of a technique for a research and development program evaluation. *Operations Research* **7** (1959), 646–669.

[12] Pleguezuelo, R. H. et al.: A note on the reasonableness of PERT hypotheses. Operations Research Letters **31** (2003).

[13] Premachandra, I. M.: An approximation of the activity duration distribution in PERT. *Computers & Operations Research* **28** (2001).

[14] Raz, T., Barnes, R., Dvir, D.: A critical look at Critical Chain project management. *Project Management Journal* **34** (2003), 24–32.

[15] Trietsch, D., Kenneth, R. B.: PERT 21: Fitting PERT/CPM for use in the 21st century. *International Journal of Project Management* **30** (2012), 490 – 502.

Sensitivity of monetary poverty measures on the setting of parameters concerning equalization of household size

Jitka Bartošová[1], Vladislav Bína[2]

Abstract. This contribution concerns the influence of parameter setting in case of the equalized household size on chosen measures of monetary poverty – poverty line, risk of poverty, depth of poverty and severity of poverty. The study of course and sensitivity of reaction is applied on the data of sample survey EU-SILC 2009 (see [1]) which provides a possibility to mutually compare the change in particular EU countries. The emphasis is primarily put on the graphical presentation of parameter changes in a model of equalized household size causing the change of national poverty lines (defined according to EU as 60 % of median of national income per consuming unit) and changes of the above mentioned relative poverty measures in case of particular EU member states. Let us clarify that poverty is a multifaceted concept related to the absence or insufficient quantity of resources that are generally regarded as indispensable for an individual or a household in a society. The contribution presents how the chosen measures of monetary poverty of the EU countries would be changed in case when usual definitions are altered.

Keywords: EU-SILC database; poverty measure; consumption unit; sensitivity.

JEL Classification: C82
AMS Classification: 62D07

1 Comparison of monetary poverty in EU

The results of measurements of monetary poverty depend not only on the definition of poverty threshold but also on the definition of equalized (i.e. mutually comparable) quantity of household incomes. The total disposable household income contained straightforwardly in the EU SILC (Statistics on Income and Living Conditions) survey data depends on the count of economically active members and pensioners living in one household. For this reason the household income itself cannot be used for the comparison of poverty or abundance of households with different counts of economically active members, respectively pensioners. The total disposable income of household (*income per household*) threats the household as a whole and for sake of monitoring of poverty or abundance it concerns only the existence of common expenditures of household. The other extreme possibility, the *income per representative*, neglects all common expenditures and considers strictly expenditures of individuals who are considered as equivalent from the aspect of monetary resources usage.

In order to obtain mutually comparable (equivalent) values of household incomes the unified measures of household size were defined, so called *consuming units* which reflect not only the count of household members but also household's structure. The transformation of incomes into the equivalent scale was performed using the definition of Organisation for Economic Co-operation and Development (OECD) – a consuming unit CU_{OECD}. This definition was later in EU modified into the variant CU_{EU}. The different parameterization of particular individuals in construction of both types of consuming units implies the change of the representative value of household income and therefore the change in evaluation of the household's poverty or abundance (see [6]). Thus the question arises concerning the sensitivity of poverty measures used in EU on the setting of weights in consuming unit definition. For this assessment of sensitivity the computational experiments were used or the monitoring of both total and partial influence of parameter changes on the change of chosen poverty measures.

1.1 Measurement of monetary poverty

In the recent past and in present the significant part of research is devoted to the question of measurement and elimination of poverty (see for example [4], [11] etc.). In literature dedicated to such issues we encounter several different views of poverty (see [8]). But there are two primary approaches – the objective and subjective one. The objective approach defines poverty by means of certain criteria concerning income or assets of a person. In

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

contrary, the methods of subjective approach investigate whether the person considers himself/herself poor and perceives the symptoms of poverty or ranks himself/herself among poor (see [5]).

Within the framework of objective approach we further distinguish the absolute and relative methodologies. The absolute methodologies define poverty using some fixed value. The relative ones define the poverty via relationship to an important characteristics (average or median income, distribution of income categories, etc.). The second concept brings information about relative poverty which is expressed by prof. Peter Townsend from London School of Economics who asserts that a person, family or population group can be considered as poor if they lack resources to obtain food, live conditions and achievements standard in communities they belong to (see [10]). In case of this definition the level of society development and prevailing circumstances are considered.

The importance of social context on determination of poverty is emphasized also in definition accepted by European commission in 1984. According to this definition, as poor can be considered a person, family or group of individuals whose resources (material, cultural and social) are so limited that they disqualify such people from minimally acceptable way of life in member states they live in. For the evaluation of poverty or abundance of household or individuals in EU member states the European commission chose so called *monetary poverty*. Among the basic comparison criteria ranks beforehand given "typical" level of income separating households (or individuals) endangered by monetary poverty from the others (see [3]). Such a line, so called *threshold of risk of monetary poverty*, is prescribed by EU on 60 % of median of national equivalent income scaled on single currency Euro in purchasing power parity. Household is thus considered as "monetary poor" if it's disposable income scaled by consuming unit (so called equalized income) lies beneath the poverty threshold.

The measures of poverty employed in EU countries stem from the class of Foster-Greer-Thorbecke poverty measures (see [2]) in general given by formula

$$P_\alpha(y,z) = \frac{1}{n}\sum_{i=1}^{q}\left(\frac{z-y_i}{z}\right)^\alpha, \tag{1}$$

where $z > 0$ is a beforehand given poverty threshold, $\mathbf{y} = (y_1, y_2, \cdots, y_n)$ is a vector of household incomes sorted by size $(y_1 \le y_2 \le \cdots \le y_q \le z)$, q is the number of households belonging to the group under the poverty threshold and n is the total count of households. Parameter α conditions the measure of sensitivity of deprivation in case of households belonging below the poverty threshold (see [7]). For $\alpha > 1$ the value of P_α begin to be distributionaly sensitive and with growing value of α grows the sensitivity on measuring the poverty of the poorest. For $\alpha \to \infty$ P_α reflects the poverty of the poorest persons (see [8]).

The most commonly used measure of monetary poverty, so called *head count index* or *risk-of-poverty-rate* can be obtained by choosing $\alpha = 0$. We arrive at formula

$$P_0(y,z) = H = \frac{q}{n}, \tag{2}$$

which gives the ratio of population with income y not greater than poverty threshold z. The main advantage of this measure is its simplicity. In contrary, the disadvantage of P_0 is its very low sensitivity on changes in the depth of poverty. If the poor person became even poorer, the value of H did not change. Within the EU the *head count index* (or *risk-of-poverty-rate*) is defined as the ratio of persons with equivalent disposable income under 60 % of median of the national disposable income. By choice of $\alpha = 1$ we obtain another measure,

$$P_1(y,z) = PG = \frac{1}{n}\sum_{i=1}^{q}\frac{z-y_i}{z}, \tag{3}$$

describing the *depth of poverty* (or *poverty gap*) which is based on the summary evaluation of poverty according to the poverty threshold. The value of PG relates to the distance of poor from the poverty threshold. Thus we obtain information about the extent of poverty. But even this measure is not sensitive enough when the "poor person" becomes "very poor". This lack of sensitivity will be removed by choice of $\alpha = 2$.

For $\alpha = 2$ we obtain so called *severity of poverty (*or *squared poverty gap)*

$$P_2(y,z) = P_2 = \frac{1}{n}\sum_{i=1}^{q}\left(\frac{z-y_i}{z}\right)^2. \tag{4}$$

The main advantage of this measure is the consideration of inequality among poor (it means that the transfer from poor to even poorer is registered by this measure). The main disadvantage is its uneasy interpretation. But despite of this fact it is considered as an appropriate tool for comparison of situation in case of the poorest.

1.2 Parametrical model of consuming unit

The value of poverty threshold and the above mentioned measures of monetary poverty is strongly influenced by the choice of consuming unit definition. This definition determines the rules for transformation of number of household members into the value (size of consuming unit) representing the modified size of household. The point of this transformation is to provide a mutual comparability of household poverty or abundance regardless of the count and age structure of household members.

The construction of consuming unit is based on an assumption that the social situation of households is dependent not only on the total amount of annual incomes of household but also on the expenditures of two types:

- common expenditures covering the functioning of household (housing expenditures, water and energies, equipment of household by durable goods, etc.);
- total expenditures on satisfying the individual needs in common household (expenditures on food and beverages, alcohol and tobacco, clothing and footwear, etc.).

Therefore the calculation of consuming units (*CU*) representing the size of household involves only one person (head of household) with the full weight (1). Weights of other members are lowered. All remaining members are for sake of parameterization of consuming unit sorted into two groups according to age. The first group contains the children of age 0 – 13 and is considered with the weight k_1. The second group comprises older children and other household members and is included in the sum with weight of k_2. For the sake of presentation of total incomes in the chosen scale, i.e. as *incomes per consuming unit* (y/CU) the following model is generally used

$$CU(k_1,k_2) = 1 + k_1 n_1 + k_2 n_2; \quad k_1, k_2 \in \langle 0;1 \rangle, \tag{5}$$

where n_1 is the count of children between 0 – 13 years and n_2 is the number of other household members.

In contemporary EU the presentation of incomes in *equivalent scale* stems from the definition of *modified consuming unit* given by formula $CU_{EU} = 1 + 0{,}3\ n_1 + 0{,}5\ n_2$. This unit originates from the modification defined by Organisation for Economic Co-operation and Development in the form of $CU_{OECD} = 1 + 0{,}5\ n_1 + 0{,}7\ n_2$.

The process of modification thus decreased both parameters (k_1 and k_2) by two tenths. This emphasized the first component (common expenditures) in comparison to the second constituent (individual expenditures of household members). In consequence, the relative growth of incomes in modified scale took place (in comparison to the original OECD scale) in case of households with higher count of members. Therefore the modification changed the perspective of the monetary poverty of households. Relatively smaller households can now easier fall under the poverty threshold. The emphasize of common expenditure component corresponds rather to the situation in countries of western Europe where the housing expenditures comprise higher percentage of total expenditures of household. Until now, in post-communist states (despite of the lasting growth of housing expenditures) the situation persists corresponding rather to the OECD scale.

The change of both coefficients into the extreme values ($k_1 = k_2 = 0$ and $k_1 = k_2 = 1$) we obtain radical change of the perspective of income situation of household and its monetary poverty. The setting of zero weights implies that consuming unit is a *household as a whole* ($CU_H = 1$), the choice of ones the value of consuming unit corresponds to the *count of household members(persons)* ($CU_I = 1 + n_1 + n_2$).

The parametrical experiments regarding the setting of scale for calculation of incomes per consuming unit are performed by change of any of parameters particularly or of both simultaneously. The change can be theoretically performed continuously in the whole domain (i.e. for $[k_1, k_2] \in \langle 0;1 \rangle \times \langle 0;1 \rangle$). The change of parameters projects into the change of results in case of monetary poverty measures. According to the fact that we observe the dependence of a continuous variable on two continuous variables (parameters k_1 and k_2), the result can be visualized using the 3D graph (as shown on the left part of Figure 1).

For sake of simplicity a easier visualization, we can choose one of the cuts in the parametrical model $CU(k_1, k_2)$ and perform the analysis of changes in monetary poverty measures only for this particular case. The aim followed during the construction of the cut is to achieve the combination of two parameters k_1 and k_2 into a single parameter k in such manner that the cut *CU(k)* goes through all distinguished values. The cut itself will pass through all the above mentioned definition of consuming unit – unit defined according to the methodology of OECD (CU_{OECD}), modified unit (CU_{EU}), but also both extreme values CU_H and CU_I. Thus we need to construct

a curve passing through four fixed points. All these requirements together with simplicity and differentiability lead us to the polynomial of degree three. Choosing

$$k_1 = k, \qquad k_2 = a_3 k^3 + a_2 k^2 + a_1 k \tag{6}$$

the curve of the cut can be written in the form

$$CU(k) = 1 + k \cdot n_1 + \left(a_3 k^3 + a_2 k^2 + a_1 k\right) \cdot n_2 ; \quad k \in \langle 0;1 \rangle . \tag{7}$$

Coefficients a_1, a_2, a_3 can be obtained as a solution of the system of three equations

$$\begin{aligned} 0{,}7 &= a_3 \cdot 0{,}5^3 + a_2 \cdot 0{,}5^2 + a_1 \cdot 0{,}5 \\ 0{,}5 &= a_3 \cdot 0{,}3^3 + a_2 \cdot 0{,}3^2 + a_1 \cdot 0{,}3 \\ 1 &= a_3 \cdot 1^3 + a_2 \cdot 1^2 + a_1 \cdot 1 \end{aligned} \tag{8}$$

The solution $a_1 = \frac{229}{105}; a_2 = -\frac{68}{35}; a_3 = \frac{16}{21}$ provides a monoparametrical model of consuming unit

$$CU(k) = 1 + k \cdot n_1 + \left(\frac{16}{21}k^3 - \frac{68}{35}k^2 + \frac{229}{105}k\right) \cdot n_2 ; \quad k \in \langle 0;1 \rangle . \tag{9}$$

2 The effect of reparametrisation on measures of poverty in EU 12 states – results and conclusions

The effect of parameter *k* on the measures of monetary poverty in EU 12 countries is visualized on Figures 1 and 2. We can see that the reaction of the poverty threshold is represented in all states by the smooth convex curves decreasing with the decreasing value of *k* (see left part of Figure 1). The poverty threshold for the household as a whole is usually close to the poverty threshold of individual multiplied twice. Though the reaction on parameter setting in consuming unit model is in all studied states very similar, the slope of curves representing the sensitivity on increase of *k* is different. The curves on the margins – for classes with maximal and minimal poverty threshold (Luxembourg and Portugal) – are sufficiently distant from the central band where remaining 10 EU-12 states are concentrated. Therefore we cannot observe any crossing with other states.

But in the central region where the curves of model are quite dense, there occur some intersecting and crossing of the curves and therefore we can observe changes in the overall ranking. We can observe that the fastest decline is in the case of increasing *k* for the poverty threshold in Great Britain. The consequence is that as a result the curve crosses sequentially two other curves and at the end it approaches the third one. At the beginning the decrease of UK from eight's to ninth's position (bellow Italy) and then the steep decrease continues bellow Spain and finally (for $k \to 1$) the poverty threshold in UK almost coincides with the poverty threshold of Greece (see left part of Figure 1 and Table 1).

state	household	rank	unit EU	rank	unit OECD	rank	person	rank
LU	30820,20	1	19334,00	1	16874,82	1	13945,20	1
IE	22546,82	2	13920,02	2	12077,90	2	9969,10	2
DK	18021,97	3	11420,64	3	10081,59	3	8337,07	3
FR	17970,00	4	11091,00	4	9603,53	4	7986,00	4
NL	17834,40	5	11076,40	5	9675,88	5	7968,00	5
BE	16521,16	6	10243,22	6	8900,99	6	7411,89	6
DE	15240,60	7	9453,00	7	8177,11	7	6809,40	7
IT	14585,40	**9**	**9175,60**	**9**	**7960,25**	**9**	**6585,90**	**9**
ES	13680,00	**10**	**8300,00**	**10**	**7200,00**	**10**	**5873,81**	**10**
UK	**14791,57**	**8**	7698,21	**8**	6481,07	**8**	5118,19	**8**
GR	11520,00	11	7003,20	11	6105,88	11	5055,00	11
PT	8647,36	12	5332,50	12	4681,59	12	3876,00	12

Table 1 Poverty threshold in EU 12 countries (Luxembourg, Ireland, Denmark, France, Nederland, Belgium, Deutschland, Italy, Spain, United Kingdom, Greece, Portugal).

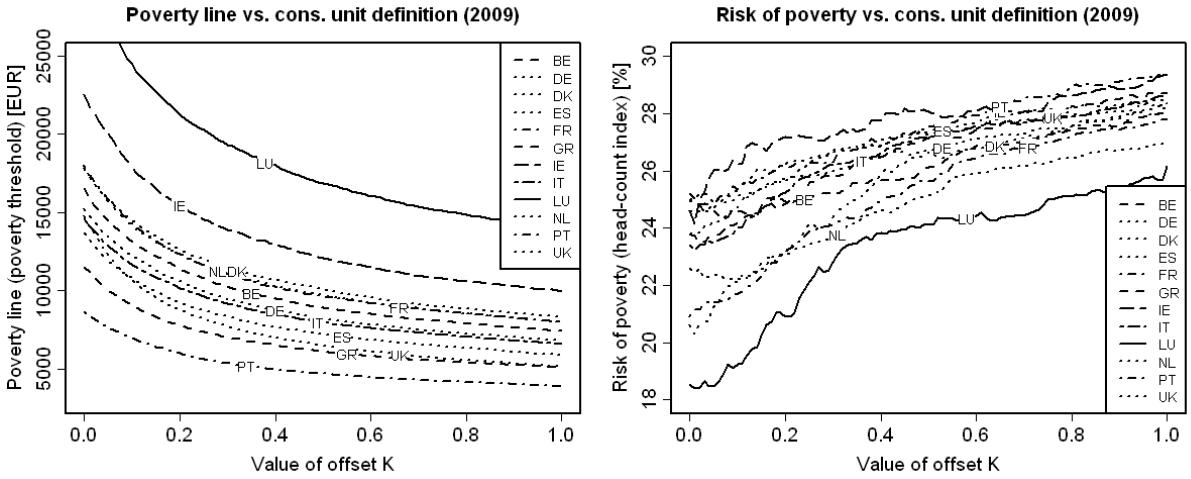


Figure 1 Dependence of the poverty threshold and risk of poverty for model of consuming unit.

The course of another important characteristic – risk of poverty – is far from being smooth (see right part of Figure 1). It is apparent that this measure responds much more sensitively on the small changes in parameter *k*. This fact together with relatively densely located curves of all studied states implies significant impact on the ranking. The curves frequently contact and cross and thus the ranking of countries changes with the change of *k* parameter. Therefore the choice of weights in the model of consuming unit (in the selected representation *CU(k)*) can influence not only the absolute value of monetary poverty in each country but also the relative location of country in the international comparison.

Similarly, though in much smaller extent behave the curves of depth and severity of monetary poverty (see Figure 2). The slope of these curves changes with *k* and for related countries frequent crossing and therefore switching of their rank occurs. Mostly these curves are rather smooth, particularly in case of severity of poverty. But always the curves are concave and increasing though (in case of severity of poverty) they are not far from being constant.

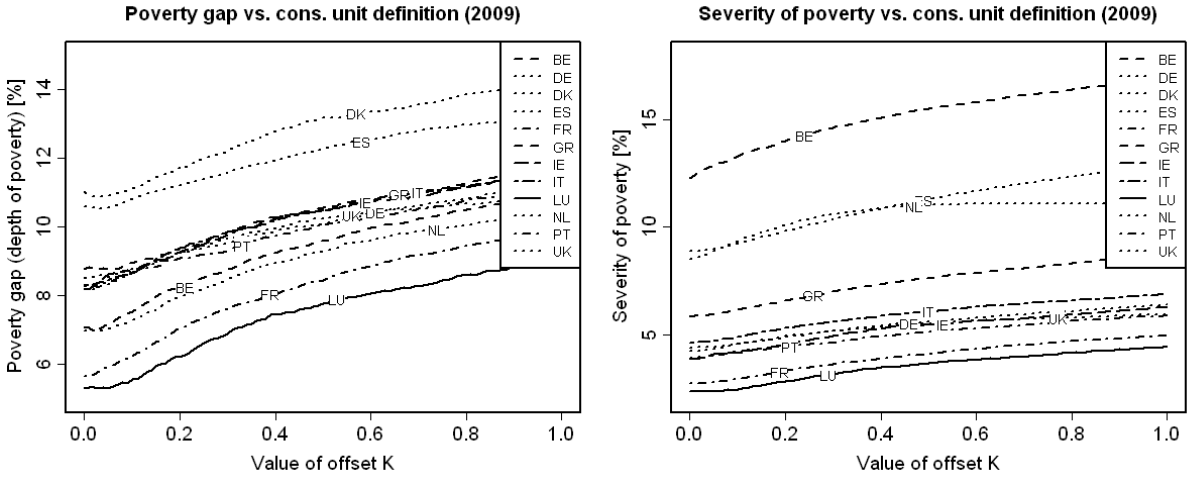


Figure 2 Curves of the poverty gap and severity of poverty for parametrised model of consuming unit.

2.1 Testing the dependence of threshold, risk, depth and severity of poverty on the setting of parameter *k* in group of EU12 states.

For the testing of statistically significant dependence of the threshold, depth and severity of poverty on the distinguished values of parameter *k* (0, 0.3, 0.5, 1) in the consuming unit model in case of EU 12 states the two-way analysis of variance was used. In case of risk of poverty the Pearson of independence was employed. For all measures except the severity of poverty ANOVA detected significant dependence (see Table 2), similarly the risk of poverty appeared dependent on the consuming unit definition (p-values approaching zero for all states).

poverty measure	factor	F-value	p-value	signif. codes
poverty threshold	CU(k); k = 0, 0.3, 0.5, 1	115.6220	$< 2.2 \cdot 10^{-16}$	***
	EU 12 countries	31.8660	$2.063 \cdot 10^{-14}$	***
depth of poverty	CU(k); k = 0, 0.3, 0.5, 1	421.6600	$< 2.2 \cdot 10^{-16}$	***
	EU 12 countries	167.9200	$< 2.2 \cdot 10^{-16}$	***
severity of poverty	CU(k); k = 0, 0.3, 0.5, 1	0.1674	0.9176000	
	EU 12 countries	55.1237	$< 2.2 \cdot 10^{-16}$	***

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 2 Results of dependence testing in case of threshold, depth and severity of poverty: two-way ANOVA.

According to the fact that in all ANOVA runs slight violation of normality assumption was detected, also a non-parametric variant (Friedman test) was used. The results appeared to be identical, therefore we can conclude that three of the measures – threshold, risk and depth of poverty – depend significantly on the choice of parameter k (k = 0; 0.3; 0.5; 1) in the model of consuming unit CU(k).

3 Conclusion

The contribution presents a sensitivity study of the response of monetary poverty measures on the change of parameters in case of date from EU SILC 2009 survey (see [1])**.** The analysis confirmed the dependence of ranking of EU 12 countries in charts mutually comparing threshold, risk, depth and severity of monetary poverty. The presented rankings are thus conditioned by the choice of equivalent scale used for the transformation of total incomes into the mutually comparable form. The choice of model parameters thus predetermines to some extent the result. Except for the severity of poverty all measures appeared to be significantly dependent on the choice of consuming unit definition. It means that even the narrower European region (particularly EU 12 group) is not compact enough since it depends on the point of view we are assessing the financial situation of the household, i.e. whether we consider relevant to emphasize during the construction of consuming unit the common expenditures of household or the expenditures on satisfying of individual needs. All computations and graphical outputs were realized using the R software (see [9]).

Acknowledgements

The article was partially supported by project IG F6/3/2012 "*Quantitative Study of the Social Situation of Juniors and Seniors*" of the Internal Grant Agency of University of Economics in Prague and by project VEGA 1/0127/11 "*Spatial Distribution of Poverty in the European Union*" of the Slovak Scientific Grant Agency.

References

[1] Eurostat. *EU-SILC User Database description (Version 2007.1 from 01–03–09)*. European Commission. Luxembourg, 2009.

[2] Foster, J., Greer, J. and Thorbecke, E. A Class of Decomposable Poverty Measures. *Econometrica* **52(3)**, 1984, 761 – 766.

[3] Goedhart, T. et al. The Poverty Line: Concept and Measurement. *The Journal of Human Resources* **12(4)**, 1977, 503 – 520.

[4] Guio, A. C. Income poverty and social exclusion in the EU25. *Statistics in Focus. Population and Social Conditions*. Eurostat, Luxembourg, 2005.

[5] Kapteyn, A. Kooreman, P. and Willemse, R. Some Methodological Issues in the Implementation of Subjective Poverty Definitions. *The Journal of Human Resources* **23(2)**, 1988, 222 – 242.

[6] Longford, N. T., and Nicodemo, C. A sensitivity analysis of poverty definitions. *IRISS Working Paper Series 2009–15*. CEPS/INSTEAD, Differdange, Luxembourg, 2009.

[7] Morduch, J. *Poverty Measures. Handbook on Poverty Statistics: Concepts, Methods and Policy Use*. New York: United Nations, Department of Economic and Social Affairs, 2005.

[8] Ravallion, M. *Poverty Comparisons: A Guide to Concepts and Methods*. Washington, DC: Světová banka, 1992.

[9] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006.

[10] Townsend, P. *Poverty in the United Kingdom: A Survey of Household Resources and Standards of Living*. Harmondsworth: Penguin Books, 1979 (Reprinted 1983).

[11] Želinský, T. Analýza chudoby na Slovensku založená na koncepte relatívnej deprivácie. *Politická ekonomie* **58(4)**, 2010, 542 – 565.

Resolved and unresolved problems in the theory of redistribution systems

Jiří Benesch[1], Hana Mihalčinová[2], Radim Valenčík[3]

Abstract. This study adopts the approach (and context) taken by J. Neumann and O. Morgenstern in describing, defining and finding solutions to simple majority game of three players and applies it to finding similar solutions in the redistribution system of three players. This system allows us to analyse situations in which the volume of what can be divided between players is determined by the way the players divide it, i.e. it is one of the examples of a non-constant sum game. As we anticipate a fully symmetric situation, we may define the term of "expected average payoff". From the term "expected average payoff" the concept of commonly acceptable equilibrium is derived. The distribution of wage at an acceptable equilibrium point is (in general) close to Nash's solution to a relevant cooperative game, which is derived from the point whose coordinates coincide with the expected average wage, yet are not completely identical. In the conclusion we outline the practical use of the model based on the definition of the redistribution system as well as a commonly acceptable equilibrium in the context of testing hypotheses concerning the objects called structures based on mutual covering-up of violation of generally accepted principles. The concept of these structures was derived from analysing games such as The Tragedy of the Commons, in the context of current problems (e.g. corruption and related issues).
Keywords: simple majority game, redistribution system, discrimination equilibrium, commonly acceptable equilibrium, structures based on mutual covering-up.

JEL Classification: C70
AMS Classification: 90C80

1 Three-person games in the classic book theory of games and economic behavior

Published for the first time as long ago as in 1944, the classic work by J. Neumann and O. Morgenstern Theory of Games and Economic Behavior [8] provides major theoretical background on the description and solution to certain types of multi-player games. We will demonstrate that Neumann's and Morgenstern's ideas can be applied to analysing three-person games played in redistribution systems. This will allow us to identify and describe the discrimination equilibrium and consequently the commonly acceptable equilibrium, which are both highly relevant for understanding people's real behaviour, including its ethical aspects. Although superseded in many respects, the work by J. Neumann and O. Morgenstern contains certain important points, which have not been considered enough in later research. Let's now focus on how the aforementioned book analyses the issue of negotiations between three persons. The basic case comes in § 21: *The Simple Majority Game of Three Persons.* The following are the most important passages and paragraphs from which they were taken: *"Each player, by a personal move, chooses the number of one of the two other players. Each one makes his choice uninformed about the choices of the two other players. … If two players have chosen each other's numbers we say that they form a* ***couple****. Clearly there will be precisely one couple, or none at all. If there is precisely one couple, then the two players who belong to it get one-half unit each, while the third (excluded) player correspondingly loses one unit. If there is no couple, then no one gets anything… Since each player makes his personal move in ignorance of those of the others, no collaboration of the players can be established during the course of the play."* [8, pp. 222-223.] As thoroughly described, the game may end up either in two players receiving ½ each and the third player -1, or in each player obtaining 0. This is one of the simplest three-person games, yet it can be extended into a more complex one. In § 21.3., the authors stress that *"the game is wholly symmetric with respect to the three players"* [8, p. 224]. This statement will prove very important. The authors are rather specific in claiming that any potential agreement among the players will always be reached outside the basic game (i.e. it would be an outcome of another game). As a follow-up, the authors take the first step and extend the basic (elementary)

[1] University of Finance and Administration, Estonská 500, 101 00 Prague 10, benesch@centrum.cz.
[2] University of Economics. W. Churchilla 4, 130 67 Prague 3. vyslouzilova@centrum.cz.
[3] University of Finance and Administration, Estonská 500, 101 00 Prague 10, valencik@seznam.cz.

model of the simple majority game of three persons (§ 22.1.2.): *"…let us now consider a game in which each coalition offers the same total return, but where the rules of the game provide for a different distribution. For the sake of simplicity, let this be the case only in the coalition of players 1 and 2, where player 1, say, is favored by an amount ε… If the couple 1,2 forms, then player 1 gets the amount ½+ε, player 2 gets the amount ½–ε ,and player 3 loses one unit. If any other couple forms (i.e. 1,3 or 2,3) then the two players which belong to it get one-half unit each while the third (excluded) player loses one unit. – What will happen in this game? – …Prima facie it may seem that player 1 has an advantage, since at least in his couple with player 2 he gets more ε than in the original, simple majority game. – However, this advantage is quite illusory. If player 1 would really insist on getting the ε in the couple with player 2, then this would have the following consequence: The couple 1,3 would never form, because the couple 1, 2 is more desirable from 1's point of view; the couple 1, 2 would never form, because the couple 2,3 is more desirable from 2's point of view; but the couple 2,3 is entirely unobstructed, since it can be brought about by a coalition of 2,3 who then need pay no attention to 1 and his special desires. Thus the couple 2,3 and no other will form; and player 1 will not get ½+ε nor even one-half unit, but he will certainly be the excluded player and lose one unit. – So any attempt of player 1 to keep his privileged position in the couple 1,2 is bound to lead to disaster for him. The best he can do is to take steps which make the couple 1,2 just as attractive for 2 as the competing couple 2,3. That is to say, he acts wisely if, in case of the formation of a couple with 2, he returns the extra ε to his partner."* [8, p. 226] This point cannot be overstressed. It explains in depth why the players in the winning coalition have to share their payoff equally. If one of them wanted more, he would find himself outside the coalition and end up losing, rather than profiting (in a zero-sum game). We will further extend the model described in the book and address a case of different amounts that can be gained at the expense of the third player if the two remaining players form a coalition. The problem is described in paragraph 22.2., entitled *Coalition of Different Strength* and can be briefly summarised as follows: Let's assume there are amounts *a, b, c* (*a* = what players 2 and 3 may get from player 1, etc.). If player 1 wanted payoff *x*, then players 2 and 3, after subtracting payoff *x,* must be left with more than or as much as players 2 and 3 would obtain from player 1 if 2 and 3 cooperated, i.e. $(c–x)+(b–x)\geq a$. This means $x\leq(–a+b+c)/2$. Thus player 1 may count upon obtaining the maximum payoff of $\alpha=(–a+b+c)/2$, and likewise players B and C may expect obtaining payoffs $\beta=(a–b+c)/2$ or $\gamma=(a+b–c)/2$. [8, p.228] Let us add the following to this brief summary: In every winning coalition, each player allied with either of the other players gets the same payoff. Or, to put it in the language of economic theory, the opportunity costs of forming any possible coalition are equal, based on the player's potential payoff in another coalition. In the next section we will look at how the initial very simple theoretic model of three-player games and their analysis can be further extended by analysis of more complex games containing other, important and real life elements.

2 Basic terms from the redistribution systems theory

In the next section we will concentrate on analyses of games played in redistribution systems. The main difference is that these games are inconstant-sum games. The redistribution systems theory aims at identifying and describing the general features of group behaviour of people in companies, workplaces, teams, institutions, organisations, etc., i.e. in places where people work together and can share the outcome of their joint performance. In redistribution system games, players need to choose (in one way or another) between their own good (i.e. coalition) and the entire system benefit. Most generally, a redistribution system can be defined as one in which the amounts shared by players depend on how players share it, with this dependence well-known and expressible (for example, by equation). The theory of redistribution systems is addressed in a monograph by P. Budinský, R. Valenčík et al. [3] One of the possible interpretations of what can be defined as a redistribution system is as follows: Assume there are three players, named A, B, C. These players perform differently; their respective performances can be expressed as e_1 (performance of player A), e_2 (player B performance), e_3 (player C performance). If each player is rewarded according to his performance, together they will achieve the highest performance $E = e_1 + e_2 + e_3$. If the reward does not correspond to the performance, the overall the performance of the system will decline; the greater the gap between the performance and the reward, the greater the decline. This can be expressed by term $\eta R(x - e_1; y - e_2; z - e_3)$, where η is a coefficient; R is the function describing the relevant dependence; x, y, z are the payoffs for individual players. We assume that the function R satisfies the distance axioms and is continuous. Payoffs for three players lie in the redistribution area based on the redistribution equation:

$$x + y + z = E - \eta R\left(x - e_1; y - e_2; z - e_3\right) \tag{1}$$

We will further assume that all points in the redistribution area are Pareto efficient, or in other words, that players took advantage of all possible Pareto improvements. [13] describes an interesting interpretation pertaining to the type of game entitled Tragedy of Commons, which is analogous to the Prisoner's Dilemma but designed for multiple players. Let's take farmers in a drought-troubled country and their limited use of water as an example. The matrix includes one of the farmers on the one side and the others on the other side. If all

farmers (both the single one and the others) adhere to their agreement and cooperate (which, in this case, means obey the agreed restrictions on the use of water), both groups would get the highest yields per hectare. Collective and unanimous breaching of the agreement (i.e. failing to cut the use of water) would lead to much lower yields on all parts. If the rules are broken by only a single farmer, his yield would be much higher, leaving virtually no impact on the others. If, by contrast, all farmers minus the one individual farmer breach the agreement, his yield will be even lower than if he joined the rest in violating the agreement. If we consider the other players to be rational and striving to maximise their own benefits, then each of them will view the situation from his own perspective and his dominant strategy will be to violate the restrictions and the final result will be disastrous. Let us note that it is the context of these tasks that provoke general aversion to models which are based on the assumption that players are rational and always aim to maximize their benefit. Consequently, it incites the conviction that the real behaviour of people can only be explained if the models are complemented with an ethical dimension, which is viewed as an exogenous element. E. Ostrom offers a different view of addressing this dilemma [9]. Based on extensive empirical results, she presents a solution in the form of self-governance. Common resources can be managed by the community, without central management. A voluntarily established community can spontaneously create an efficient management of common resources. In other words, a voluntarily established community can protect common ownership, distribute the yields of that ownership among its members, and eliminate unentitled parties. Rather than creating mathematical models, by analysing extensive empirical material from various parts of the world E. Ostrom explores social institutions and applies an evolutionary view. We will demonstrate the existence of two types of equilibriums in redistribution systems, which we disclosed here, pertain to the aforementioned issue. Let us look at a three-person game where the players use precious resource and are allowed to distribute it in a manner consistent with the Tragedy of Commons type of game. We will consider a more complex case, when the individual players (farmers) farm under different conditions. Here each use of additional unit of water results in a different payoff for each of the players. From a microeconomic point of view, the maximum common income can be achieved when the marginal income from the last unit of water to be used by any of the players equals the marginal income from the last unit of water of any other player. Now suppose that players share the water at 6:4:2 ratio. If they diverge from that ratio, their common income will be lower – the more they diverge from that water distribution ratio, the lower their income. We may also accept another assumption, which is intuitively obvious and does not narrow the universality of our problem: if they share the water at 6:4:2, their payoffs (coming from income) will be distributed accordingly, and moreover, the farmers' incomes are proportionate to the allocations of water (this proportion does not necessarily have to be linear). Here the general rule of the relationship between the payoffs and the distribution of water will be as follows: what the farmers can distribute among themselves equals the maximum of what they would be able to distribute (i.e. 12) minus the decline in common income, which results from the fact that the players will not share their income according to optimal proportion based on the equality of marginal income (i.e. the proportion of 6:4:2). This dependence is described by the redistribution equation (1). Function R can be, for example, the generally used Euclidean distance (for the ratio of 6:4:2): R[(x-6); (y-4); (z-2)] = √[(x-6)2+(y-4)2+(z-2)2]; Manhattan distance as the sum of the absolute values of the differences in performance and payoffs of the individual players: R[(x-6); (y-4); (z-2)] = |x-6|+|y-4|+|z-2|; Chebyshev distance, which always selects that difference from among the differences in the performance and payoffs of the individual players that appertains to the player with the greatest divergence: R[(x-6); (y-4); (z-2)] = max[(x-6); (y-4); (z-2)]. [3, pp. 51-73] Further on we will only deal with such areas which only contain points representing Pareto optimal situations (this applies to all of those mentioned above), and will search for solutions on that areas.

3 Discrimination equilibrium and commonly acceptable equilibrium

Let us try to define a very simple game (similar to those we mentioned for zero-sum games from the book by J. Neumann and O. Morgenstern). If two players form a coalition, they can distribute among themselves everything the third player would get. The third player obtains the lowest possible payoff (say 0 in our case; however, there are situations in which the player could receive more and even less than 0). The players who have created a two-person coalition will subsequently distribute payoffs in certain proportion among themselves. The major difference between this case and the one described by J. Neumann and O. Morgenstern is that the amount distributed among the players is not constant. Let us consider lines in which the payoff of one of the players equals 0 to be the so called discrimination lines, i.e. the lines of full discrimination of one player by the other two players. They are formed by points shared by the redistribution area with sides determined by two coordinates of x, y, z (Figure 1). Their course is determined by equations:

$$y + z = E - \eta R(0; y - e_2; z - e_3) \tag{2}$$

$$x + z = E - \eta R(x - e_1; 0; z - e_3) \tag{3}$$

$$x + y = E - \eta R(x - e_1; y - e_2; 0) \tag{4}$$

The question is what payoff to allocate to each of the players if they form a two-person coalition (hereinafter referred to as the 'coalition') and if they fully discriminate the third player. Let us recall what happens in a zero-sum game: First, we allocated the value corresponding to the ½:½ distribution. Then we admitted the possibility that one of the players might claim more, and proved that, unless he wished to become discriminated, the player could not claim more. Afterwards, we addressed a situation in which what is distributed is determined by whoever forms the coalition, and laid down a distribution rule: in creating a coalition with a player, every player must claim exactly what he would have claimed if he created coalition with another player. The third requirement can be also met in our more general case game of three persons. Just consider the equations that describe the course of the discrimination lines to be a system of three equations with three unknowns. If the function R satisfies the distance axioms, it is continuous and all points in the redistribution area defined by that function are Pareto-optimal. The result of this (in non-negative values) is three points, to which the following applies: The payoff for player A (i.e. x) would be the same in the coalition with player B or C. The payoff of player B (i.e. y) would be the same in coalition with player A or C. The payoff of player C (i.e. z) would be the same in the coalition with player A or B. Let us call the points located on the discrimination lines and complying with the above system of equations the points of discrimination equilibrium (Figure 1).

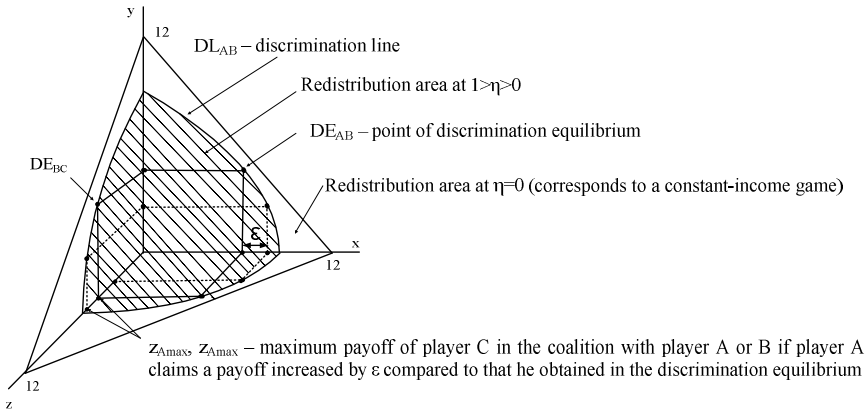


Figure 1 Redistribution area with an indication of discrimination equilibrium

Let the payoff values of players A, B, C, if they are in a coalition that discriminates the third player and if they find themselves in the points of discrimination equilibrium, be called x_d, y_d, z_d and let us suppose that the player outside the coalition will obtain a payoff equal to 0. Let us now introduce the concept of a "better" point in the sense of Neumann-Morgenstern , i.e. point x is better than point y only if the payoff for either of the two players is greater in point x than y. The points of discrimination equilibrium constitute an internally and externally stable set (for a proof, see [3 pp. 49-50]), which complies with [8, § 4.5.3, p. 40, § 32.2., pp. 282-288).]) and can be considered a solution to one type of games played in the set corresponding to the redistribution area. Let us look at another interesting feature of the points of discrimination equilibrium. If player A in a coalition with player B required a payoff by some $\varepsilon > 0$ greater than would be adequate to the discrimination equilibrium, the coalition of players A and B would never be formed – with the reasons being the same as stated in the above-cited passage [8]. Each player may require a payoff for himself only increased by $\varepsilon = 0$. If we considered a more complex game with multiple rounds of negotiations, and wished to express it explicitly (which exceeds the possibilities of our paper), player C would receive more from a coalition with player B than with player A. The reason is that his maximum payoff from a coalition with player A would be z_{Amax} while his possible payoff with player B would be z_{Bmax}, which is necessarily higher because player A claims more than would be adequate to achieve the discrimination equilibrium (Figure 1). Let us reiterate that the positions of all three players in our basic model are fully symmetric. They feel no particular aversion or preference for each other, i.e. the final composition of the coalitions is not influenced from the 'outside'; it is merely the outcome of what happens in the relevant redistribution system. We ought to make a brief note on what was said above in terms of the practical importance of the models we deal with. If we know what happens in the basic model, which is not subject to any external effects, and compare this to what happens in real-life systems, which we model by means of redistribution systems, we can find various deviations. By analysing these deviations we may subsequently be able to find the external factors (including the hidden ones) influencing the system, which we would not be able to find without the model. This is one of the practical contributions of our activities. However, in this study we explore the basic model which does not reckon with any external effects. What payoff may each player expect? Either the one he gets if he forms a coalition with one of the other two players, or the lowest payoff, which the other players will allocate to him if he is in the discrimination position, i.e. 0. Here it is relevant to state that the

player may count upon *the expected average payoff*, which equals $2x_d/3$ for player A, $2y_d/3$ for player B, and $2z_d/3$ for player C (Figure 2). The area marked by the lines of the expected average payoffs has an important feature. Each point inside of it stands for *a Pareto improvement for each player vis-à-vis his average payoff.* This means that instead of an uncertain payoff, though it may be greater than the average expected payoff in two out of three cases but only minimal (equal to 0) in one case, a rationally behaving player is probably going to prefer reaching an agreement on the distribution of payoffs, so that the distribution is consistent with a point inside the considered area. (For the sake of simplicity, we do not consider a decline in the marginal utility of the payoff achieved; if we did, the area of Pareto improvements would be even greater.) Our results are important and may be interpreted in various ways. We can for example say that due to its mathematical foundations reality offers the players better prospects than if they attempt to achieve such results by means of discrimination, or that the mathematical foundations of our reality contain a sort of fairness, our idea of morality, etc. Let us admit, however, that these interpretations use a great deal of fiction. We should also note that the Pareto improvements of the expected average payoffs only exist in inconstant-sum games and games with more than two persons and are not applicable to constant-sum or zero-sum games. However, we can continue to elaborate this and ask the following questions: Which of the points in the area of Pareto improvements is 'the right one', i.e. the one on which the players will agree? Is there any? And if so, can players agree that their payoffs shall be distributed in compliance with it? Issues of this sort are addressed in [1], [2], [5], [6], [7], [10], [11], [12]. One of the possible and convenient answers is Nash's solution to the cooperative game at the set corresponding to the redistribution area and to points defined by $(2x_d/3; 2y_d/3; 2z_d/3)$ coordinates. In these circumstances Nash's solution brings Pareto improvement which is better than the expected average payoffs. Yet another solution can be taken into consideration, one that we find closer to real life situations. This solution is shown in Figure 2.

Figure 2 will help us answer those questions.

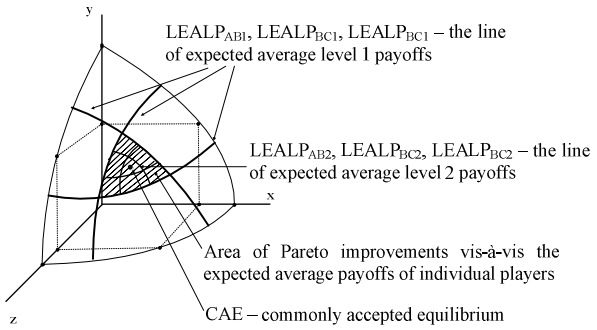


Figure 2 Redistribution area with an indication of the lines of expected average payoffs

Bearing in mind the features/characteristics of the basic game (i.e. a game analogous to the simple majority game) we can imagine that the players will realise that they are better off reaching a mutual agreement. Nevertheless, as concerns the original game, this realisation, as well as the effort for an agreement, if any, is yet another game. Its relationship to the original game is only based on the fact that the exploration of certain features of the basic game has enabled us to unveil the area of Pareto improvements in the expected average payoffs, the possibility of a joint agreement and the issue of defining a game in which the players will be choosing one of the points of those improvements. Each of the players might attempt to form a coalition with any of the other two players in order to achieve the maximum improvement even if all players are already choosing points inside the area of the Pareto improvements of expected average payoffs. The situation would reoccur (compare with [11] for two players), as even the area marked by the expected average payoff lines shares certain features with the redistribution area. It would be possible to define and calculate the points of level 2 discrimination equilibrium, to derive the lines of expected average level 2 payoffs from them, and consequently the area of the level 2 Pareto improvements. Then we could proceed to discrimination equilibriums, the lines of expected average payoffs, the area of level 3 Pareto improvements, etc. The situation resembles a matryoshka doll. Each game which enables definition of an area of Pareto improvements by certain level accommodates another game. These areas gradually become smaller and eventually clearly define one point, *the point of commonly acceptable equilibrium.* Its coordinates correspond to such payoffs the players achieve if they reach a commonly acceptable equilibrium.

4 Issues for further research. Practical application of the research. Conclusion.

It is very probable (and this can be viewed as a hypothesis) that a point of commonly acceptable equilibrium may also be reached by other procedures, e.g. by finding an intersection point of the lines derived from discrimination

equilibriums. When searching for the point of commonly acceptable equilibrium, we change parameter d_x, d_y, d_z, (the smallest payoff each player has to obtain) for each player (so far, we have considered that the value of a discriminated player's payoff is 0) from the initial value (i.e. equal to 0) to the maximum possible value (i.e. the absolutely greatest payoff the player can achieve), without changing the values of this parameter for the other two players. Three lines derived from the points of discrimination equilibria will be achieved on the redistribution area. Having established this basis, we can lay down the following two hypotheses: 1. If we change parameter d_i, the lines derived from discrimination equilibria intersect in a single point. 2. This point is identical to the point of commonly acceptable equilibrium we have achieved by the above-described negotiations. Further evidence will be required to prove whether this is the same point as the one we found as a result of the 'matryoshka doll' procedure. Several areas are to be addressed by future research, such as defining various types of games played in redistribution system, establishing rules for the players to achieve a commonly acceptable equilibrium through negotiations and thus defining the types of games in which players will fail to achieve a commonly acceptable equilibrium.

For the usual cases of the redistribution area the point of commonly acceptable equilibrium is rather close but not quite identical with the point defined by Nash. No matter how minuscule and insignificant this difference might seem to the players in real-life systems, which share some features with redistribution systems, there are certain existing social objects for which the significance of the above-described seems to be crucial. We called these objects *structures based on mutual covering-up of violation of generally accepted principles*. In the above-mentioned game "Tragedy of Commons" these situations develop when one of the players is caught by the second player when violating their agreement and, as a consequence, is blackmailed. Blackmailing than helps such structures to develop, grow and intertwine with existing social organizations or institutions. Resilience and vitality of these structures can only thrive if the structures manage to develop an inner system with features of a redistribution system in which the players are able to achieve what we called commonly acceptable equilibrium in the three-player game. This is a promising way to detect problems related to corruption and related phenomena.

Acknowledgements

This research is supported by a Czech Science Foundation (The Theory of Redistribution Systems, No. 402/09/ 0096).

References

[1] Aumann, R. J., and Hart. S.: *Handbook of Game Theory with Economic Applications.* Vol. 1-2, Elsevier, North-Holland, 2002.
[2] Binmore, K.: Bargaining and Coalitions. In: *Game Theoretic Models of Bargaining* (Roth, A. ed.). Cambridge University Press, Cambridge, 1985.
[3] Budinský. P., and Valenčík, R.: *Teorie her a redistribuční systémy*. VŠFS-EUPRESS, Praha, 2010.
[4] Budinský, P., and Valenčík, R.: *Game Theory (Redistribution and Contextual Games) as a Tool for Human Behaviour Decoding*. VŠFS, Prague 2011.
[5] Harsanyi, J.: *Rational Behaviour and Bargaining Equilibrium in Games and Social Situations.* Cambridge University Press, Cambridge, 1977.
[6] Kalai, E., and Smorodinsky, M. .: Other Solutions to Nash's Bargaining Problem. *Econometrica* **43** (1975), 513-518.
[7] Moulin, H.: *Bargaining and Non-Cooperative Implementation*. Ecole Polytechnique (Laboratoire d'Econometrie) discussion paper A239 0282, 1982.
[8] Neumann, J., and Morgenstern, O.: *Theory of Games and Economic Behavior.* Princeton University Press, Princeton 1953.
[9] Ostrom, E.: *Governing the commons: the evolution of institutions for collective action.* Cambridge University Press, New York 1990.
[10] Peleg, B., and Sudhölter, P.: *Introduction to the Theory of Cooperative Games*. Kluwer Academic Publishers, Boston, Mass, 2003.
[11] Reiffa, H.: *Arbitration Schemes for Generalized Two-person Games.* University of Michigan, 1952.
[12] Thomson, W., and Lensberg T.: Guarantee Structures for Problems of Fair Division. *Mathematical Social Sciences* **4** (1983), 205 -18.

The investment decision making under uncertainty

Adam Borovička[1]

Abstract. The article deals with the decision making process in the field of capital market, concretely with open shares funds. At the beginning we choose the set of particular shares funds for future investment. The offer of funds of Investment company Česká spořitelna is accepted by potential investor. For cut-down of extensive set of shares funds the multiple criteria evaluation method is applied. This method is based on a measurement of distance from basal and ideal alternative. Under the stochastic character of evaluation several ranking scenarios are created. The final order is stated by the assignment problem. After the fund set reduction we should select some of interactive multiple objective programming methods in order to make the final investment portfolio. In the decision making process, we take into account the DM fuzzy preferences in the sense desired values of objective functions. The applied method also includes stochastic elements leading to scenario analysis. In the end the "optimal" investment portfolio is designed.

Keywords: decision making, fuzzy set, uncertainty

JEL Classification: C61, G11
AMS Classification: 90-08, 91B28

1 Introduction

During the investment decision making process we meet several situations harbouring an uncertainty, e. g. in making investor preferences or prices development. We take into account all these matters in order to make problem more real.

Concretely we are deciding to invest in open shares fund offered by Investment company Česká spořitelna. Firstly we reduce the whole voluminous set of shares funds by the help of multiple criteria evaluation method making provision for uncertain decision maker preferences, stochastic character of some evaluative criteria as well. Then we can make "optimal" portfolio of shares funds thanks interactive multiple objective programming method also respecting fuzzy preferences and stochastic procedure. The stochastic character is represented by the presence of random variables in the investment decision making. The uncertainty in the investor preferences is also expressed via fuzzy sets.

The goal is to undergo the investment decision making with reference to all elements of uncertainty described above. For that, we will apply some principles of mathematical programming.

2 Investment situation

Imagine some potential investor who decided to insert some money into shares funds from Investment company Česká spořitelna. He chooses from four groups - *money-market funds*, *mixed funds*, *bond funds* and *stock funds* as the following table closely shows (Table 1):

Money-market funds	Mixed funds	Bond funds	Stock funds
Sporoinvest	Osobní portfolio 4	Sporobond	Sporotrend
	Plus	Trendbond	Global Stocks
	Fond řízených výnosů	Bondinvest	Top Stocks
	Konzervativní Mix	Korporátní dluhopisový	
	Vyvážený Mix	High Yield dluhopisový	
	Dynamický Mix		
	Akciový Mix		

Table 1 List of shares funds offered by Investment company Česká spořitelna [7]

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

The investor follows two criteria – *return* and *risk*. Further *costs*[2] and *Sharpe ratio*[3] also play a big role in the process. The criterion return is stated as "important" and risk as "very important". The highest possible level of costs is stated as 2 % with tolerance 0.5 %, the lowest level of Sharpe ratio 0.2 % with 0.2 % tolerance to cover at least the risk-free yield rate. Further he requires the minimum share of one shares fund 5 % and the maximum level 50 % with the view of portfolio diversification. Finally only one fund from each group has to be in final investment portfolio.

3 Multiple criteria evaluation method

As the investor wants to have only one representative from each group of shares funds in portfolio, we use some multiple criteria evaluation method in order to choose the most suitable candidate in each group. For this purpose, the following method is proposed.

This approach accepts stochastic character presented by *return* as a random variable and fuzzy preferences about the weights of criteria. Let's describe the method in these several steps.

Step 1: Given the matrix of evaluation $Y = (y_{ij})$, where y_{ij} $(i = 1,2,...,p;\ j = 1,2,...,k)$ represents valuation of i^{th} variant by j^{th} criterion. The input information in the form of weight vector $v = (v_1, v_2, ..., v_k)$ can be determined as stochastic, where v_j is a random variable with uniform probability distribution $R(a,b)$, y_{ij} for some j as well (like other assumption). For allowance, all minimizing criteria are transformed to maximizing form as follows

$$q_{ij} = \max_i (y_{ij}) - y_{ij} \quad \forall j \text{ (min)} \qquad q_{ij} = y_{ij} \quad \forall j \text{ (max)}.$$

Normalize the previous values in the following formula

$$r_{ij} = \frac{q_{ij}}{g_j} \quad i = 1,2,...,p,\ j = 1,2,...,k, \qquad \text{where} \quad g_j = \max_i (q_{ij}) \text{ for } j = 1,2,...,k.$$

Step 2: Compute the ideal alternative $h = (h_1, h_2, ..., h_k)$, where

$$h_j = \max_i (r_{ij}) \quad j = 1,2,...,k,$$

and basal alternative $d = (d_1, d_2, ..., d_k)$ with a component

$$d_j = \max_i (r_{ij}) \quad j = 1,2,...,k.$$

Now we can come up to creation of distance from ideal (s_i^u), or basal (s_i^l) variant as Euclidean metric [3]

$$s_i^u = \sqrt{\sum_{j=1}^{k} v_j (h_j - r_{ij})^2} \quad i = 1,2,...,p \qquad s_i^l = \sqrt{\sum_{j=1}^{k} v_j (r_{ij} - d_j)^2} \quad i = 1,2,...,p.$$

Step 3: As we can see in [6], the shortest distance to the ideal variant does not guarantee the longest distance to the basal one in the case of Euclidean matric, so it is meaningful to calculate the relative indicator of distances from ideal alternative

$$f_i = \frac{s_i^u}{s_i^u + s_i^l} \in \langle 0,1 \rangle \quad i = 1,2,...,p.$$

The alternative order is made in accordance with ascending values of the prior indicator.

Step 4: In the spirit of stochastic character we get several scenarios of ranking. For the final order, a principle of the assignment problem[4] is applied. Formulate the matrix $T = (t_{ij})$, where $t_{ij} (i, j = 1,2,...,p)$ represents the

[2] We take into account average monthly returns from 1st April 2009 to 1st December 2011. This period must be cut for mixed shares funds *Osobní portfolio 4* and *Plus* because of their later foundation. The risk is stated as a standard deviation of fund return. The costs include the entry fees.

[3] *Sharpe ratio* measures the efficiency of particular investment instrument with regard to its riskiness [4].

[4] The basic idea of assignment problem can be described as an mutual assignment of elements from two stated sets that should be as efficient as possible, it means with minimum costs, maximum returns etc. [8]

number expressing how many times j^{th} alternative is placed on i^{th} position. In other words t_{ij} characterizes the fitness of assignment. Thus the assignment problem will be in the following form

$$z = \sum_{i=1}^{p}\sum_{j=1}^{p} t_{ij}x_{ij} \to \max$$

$$\sum_{i=1}^{p} x_{ij} = 1 \quad j = 1,2,...,p \qquad \sum_{j=1}^{p} x_{ij} = 1 \quad i = 1,2,...,p$$

$$x_{ij} \in \{0,1\} \quad i, j = 1,2,...,p,$$

where x_{ij} takes 1 if the alternative j is placed on i^{th} position, otherwise equals 0. As we try to reach the assignment as efficient as possible in the sense of values t_{ij}, the objective function z is maximized.

4 Interactive multiple objective programming method

After reduction of all groups, we come up to final portfolio creation by the help of interactive multiple objective programming method. This method takes into account the stochastic character in the shape of random quantities in the process and also uncertain investor preferences expressed by fuzzy sets.

Step 1: Firstly define all criteria and constraints as a complex problem [5]

$$\begin{aligned} &[f_1(x_1,x_2,...,x_n),...,f_k(x_1,x_2,...,x_n)] \to \text{"max"} \\ &g_i(x_1,x_2,...,x_n) \geq \tilde{b}_i \quad i = 1,2,...,m_1 \\ &g_i(x_1,x_2,...,x_n) \leq \tilde{b}_i \quad i = m_1+1,2,...,m_2 \\ &g_i(x_1,x_2,...,x_n) = \tilde{b}_i \quad i = m_2+1,2,...,m \\ &p_i(x_1,x_2,...,x_n)R_iq_i \quad i = 1,2,...,r \end{aligned} \quad \to \quad \begin{aligned} &[f_1(x_1,x_2,...,x_n),...,f_k(x_1,x_2,...,x_n)] \to \text{"max"} \\ &g_i(x_1,...,x_n) \geq b_i - b_i^* \quad i = 1,2,...,m_1 \\ &g_i(x_1,x_2,...,x_n) \leq b_i + b_i^* \quad i = m_1+1,2,...,m_2 \\ &g_i(x_1,x_2,...,x_n) \geq b_i - b_i^l \quad i = m_2+1,2,...,m \\ &g_i(x_1,x_2,...,x_n) \leq b_i + b_i^u \quad i = m_2+1,2,...,m \\ &p_i(x_1,x_2,...,x_n)R_iq_i \quad i = 1,2,...,r, \end{aligned}$$

where $f_l(l = 1,2,...,k)$ expresses l^{th} objective function (criterion), $x_j(j = 1,2,...,n)$ represents j^{th} unknown variable, $g_i(i = 1,2,...,m)$ is the left side and $\tilde{b}_i(i = 1,2,...,m)$ is the right side of the i^{th} limit. The values $\tilde{b}_i$ show an uncertainty, so decision maker (DM) does not determines the strict demands, but only b_i level with tolerance $b_i^*(i = 1,2,...,m_2)$, or b_i^l and $b_i^u(i = m_2+1,...,m)$ according to constraint type. The symbol $p_i(i = 1,2,...,r)$ is the left side, $R_i(i = 1,2,...,r)$ a relational sign, and $q_i(i = 1,2,...,r)$ represents the right side of the i^{th} limit with no fuzzy elements. This fact is shown in the model on the right side brightly quantifying the vague requirements in extreme tolerance concept. It is desirable to reach as high as possible values of all objective functions ("max").

Step 2: Set the lower L_l and upper U_l bound for the l^{th} objective. To calculate these bounds of all objective functions we first solve the following sub problems for each i^{th} objective function of minimizing or maximizing character.

$$\begin{aligned} &z_l = f_l(x_1,x_2,...,x_n) \to \max(\min) \\ &g_i(x_1,...,x_n) \geq b_i \quad i = 1,2,...,m_1 \\ &g_i(x_1,x_2,...,x_n) \leq b_i \quad i = m_1+1,2,...,m_2 \\ &g_i(x_1,x_2,...,x_n) = b_i \quad i = m_2+1,2,...,m \\ &p_i(x_1,x_2,...,x_n)R_iq_i \quad i = 1,2,...,r \end{aligned} \qquad \begin{aligned} &z_l = f_l(x_1,x_2,...,x_n) \to \max(\min) \\ &g_i(x_1,...,x_n) \geq b_i - b_i^* \quad i = 1,2,...,m_1 \\ &g_i(x_1,x_2,...,x_n) \leq b_i + b_i^* \quad i = m_1+1,2,...,m_2 \\ &g_i(x_1,x_2,...,x_n) \geq b_i - b_i^l \quad i = m_2+1,2,...,m \\ &g_i(x_1,x_2,...,x_n) \leq b_i + b_i^u \quad i = m_2+1,2,...,m \\ &p_i(x_1,x_2,...,x_n)R_iq_i \quad i = 1,2,...,r \end{aligned}$$

Identify the optimal solution of first model as x_{1j}^o, or the second one $x_{2j}^o(j = 1,2,...,k)$ with the values of objective functions $z_l^o(x_{1j}^0)$, or $z_l^o(x_{2j}^0)$ for $j,l = 1,2,...,k$. Then the lower (L_l) and upper (U_l) bounds of l^{th} objective function are calculated as follows

$$L_l = \min\{z_l^o(x_{1j}^0), z_l^o(x_{2j}^0)\} \qquad U_l = \max\{z_l^o(x_{1j}^0), z_l^o(x_{2j}^0)\}.$$

When the aspiration levels for each objective are obtained, we can form a fuzzy model where find $x_j\,(j=1,2,...,n)$ so as to satisfy

$$\begin{aligned} z_l &\tilde{\leq} L_l \quad \forall l(\min) \\ z_l &\tilde{\geq} U_l \quad \forall l(\max) \\ g_i(x_1,...,x_n) &\tilde{\geq} b_i \quad i=1,2,...,m_1 \\ g_i(x_1,...,x_n) &\tilde{\leq} b_i \quad i=m_1+1,...,m_2 \\ g_i(x_1,...,x_n) &\tilde{=} b_i \quad i=m_2+1,...,m \\ p_i(x_1,x_2,...,x_n) &R_i q_i \quad i=1,2,...,r. \end{aligned} \tag{1}$$

The membership functions[5] for fuzzy constraints of (1) are defined as a form of triangular fuzzy number.

According to [5], for l^{th} constraints (G_l) accordance with minimizing objective function, or maximizing one

$$\mu_{G_l}(z_l)=\begin{cases} 1 & z_l \leq L_l \\ \frac{U_l-z_l}{U_l-L_l} & L_l \leq z_l \leq U_l \\ 0 & z_l > U_l \end{cases} \qquad \mu_{G_l}(z_l)=\begin{cases} 1 & z_l \geq U_l \\ \frac{z_l-L_l}{U_l-L_l} & L_l \leq z_l \leq U_l \\ 0 & z_l < L_l \end{cases}.$$

For i^{th} constraint (E_i), where $i=1,2,...,m_1$, $i=m_1+1,...,m_2$, or $i=m_2+1,...,m$ in agreement with the type of limit, we can write the membership function in the mentioned order

$$\mu_{E_i}(b_i)=\begin{cases} 1 & g_i(\vec{x}) \geq b_i \\ \frac{g_i(\vec{x})-b_i+b_i^o}{b_i^o} & b_i-b_i^o \leq g_i(\vec{x}) \leq b_i \\ 0 & g_i(\vec{x}) < b_i-b_i^o \end{cases} \qquad \mu_{E_i}(b_i)=\begin{cases} 1 & g_i(\vec{x}) \leq b_i \\ \frac{b_i+b_i^o-g_i(\vec{x})}{b_i^o} & b_i \leq g_i(\vec{x}) \leq b_i+b_i^o \\ 0 & g_i(\vec{x}) > b_i+b_i^o \end{cases}$$

$$\mu_{E_i}(b_i)=\begin{cases} 0 & g_i(x) < b_i-b_i^l \\ \frac{g_i(x)+b_i^l-b_i}{b_i^l} & b_i-b_i^l \leq g_i(x) \leq b_i \\ \frac{b_i^u+b_i-g_i(x)}{b_i^r} & b_i \leq g_i(x) \leq b_i+b_i^u \\ 0 & g_i(x) > b_i+b_i^u \end{cases}$$

According to [1] the fuzzy decision is represented by fuzzy set $A=G_1\cap...\cap G_k\cap E_1\cap...\cap E_m\cap X$, where X is (non-fuzzy) set of feasible solutions of initial problem, thus $X=\{x\in R^n, p_i(x)R_iq_i, i=1,2,...,r\}$. The optimal solution $x^*\in X$ has the maximum value of membership function $\mu_A=\min_{i,l}(\mu_{G_l}(z_l),\mu_{E_i}(b_i))$. On the basis of [1], the optimal solution can be obtained via the problem of linear programming written as follows

$$\begin{aligned} &\lambda \to \max \\ z_l+\lambda(U_l-L_l) &\leq U_l \quad \forall l(\min) \\ z_l-\lambda(U_l-L_l) &\geq L_l \quad \forall l(\max) \\ g_i(x_1,...,x_n)-\lambda b_i^o &\geq b_i-b_i^o \quad i=1,2,...,m_1 \\ g_i(x_1,...,x_n)+\lambda b_i^o &\leq b_i+b_i^o \quad i=m_1+1,...,m_2 \\ g_i(x_1,...,x_n)-\lambda b_i^o &\geq b_i-b_i^l \quad i=m_2+1,...,m \\ g_i(x_1,...,x_n)+\lambda b_i^o &\leq b_i+b_i^u \quad i=m_2+1,...,m \\ p_i(x_1,...,x_n)&R_iq_i \quad i=1,2,...,r \\ 0 \leq \lambda &\leq 1, \end{aligned}$$

where $\lambda=\min_{i,l}\{\mu_{G_l}(z_l),\mu_{E_i}(b_i)\}$.

[5] In [9], the membership function measures the degree of set membership. It takes the value from interval $\langle 0,1\rangle$. The higher value denotes the higher degree of set membership.

In the case of fuzzy weights of criteria, the optimal solution $x^* \in X$ has the maximum value of membership function $\mu_A = \min_{i,l}\{\mu_{W_l}(\mu_{G_l}(z_l)), \mu_{E_i}(b_i)\}$, where μ_{W_l} represents the membership functions describing the fuzzy decision maker preferences about the criteria.

Step 3: Thanks the random variables in the process, we obtain several solution scenarios. Under the Monte Carlo concept [see more 2], the solution with the maximum value of objective function (λ) is chosen for further interactive procedure. When the current solution is acceptable, the process is finished. If not, the decision maker (investor) has some demands for solution (portfolio) improvement that can have fuzzy character, so some additional constraints will be included in the model. We select the criteria which should be made better, then new constraints are as follows

$$z_l \geq z_l^c + \Delta z_l \quad \forall l(\max) \qquad z_l \leq z_l^c - \Delta z_l \quad \forall l(\min),$$

where $\Delta z_l (l = 1,2,...k)$ expresses the desired minimal betterment of l^{th} criterion and $z_l^c (l = 1,2,...,k)$ is the current value of l^{th} objective function. Under these conditions the solution can be infeasible. Then DM has to shrink his demands to find it. Otherwise values of some criteria must be sacrificed. The DM accepts decrease maximizing, or minimizing criterion in value $\Delta^{\max}$ with tolerance $\overline{\Delta}^{\max}$, or $\Delta^{\min}$ with tolerance $\overline{\Delta}^{\min}$, thus

$$\forall l(\max) \quad z_l \tilde{\geq} z_l^c - \Delta^{\max} \quad \rightarrow \quad z_l^c - z_l \tilde{\leq} \Delta^{\max} \quad \rightarrow \quad z_l^c - z_l \leq \Delta^{\max} + \overline{\Delta}^{\max} \text{ (with extreme tolerance)}$$

$$\forall l(\min) \quad z_l \tilde{\leq} z_l^c + \Delta^{\min} \quad \rightarrow \quad z_l - z_l^c \tilde{\leq} \Delta^{\min} \quad \rightarrow \quad z_l - z_l^c \leq \Delta^{\min} + \overline{\Delta}^{\min} \text{ (with extreme tolerance).}$$

Now the membership function for new preference constraints may be declared as

$$\mu_B(\Delta^{\max}) = \begin{cases} 1 & z_l^c - z_l \leq \Delta^{\max} \\ \frac{\Delta^{\max} + \overline{\Delta}^{\max} - (z_l^c - z_l)}{\overline{\Delta}^{\max}} & \Delta^{\max} \leq z_l^c - z_l \leq \Delta^{\max} + \overline{\Delta}^{\max} \\ 0 & z_l^c - z_l > \Delta^{\max} + \overline{\Delta}^{\max} \end{cases}, \mu_B(\Delta^{\min}) = \begin{cases} 1 & z_l - z_l^c \leq \Delta^{\min} \\ \frac{\Delta^{\min} + \overline{\Delta}^{\min} - (z_l - z_l^c)}{\overline{\Delta}^{\min}} & \Delta^{\min} \leq z_l - z_l^c \leq \Delta^{\min} + \overline{\Delta}^{\min} \\ 0 & z_l - z_l^c > \Delta^{\min} + \overline{\Delta}^{\min} \end{cases}.$$

So we must add particular constraints representing fuzzy preferences in the final model in the following form

$$z_l^c - z_l + \lambda\overline{\Delta}^{\max} \leq \Delta^{\max} + \overline{\Delta}^{\max} \qquad z_l - z_l^c + \lambda\overline{\Delta}^{\min} \leq \Delta^{\min} + \overline{\Delta}^{\min}.$$

The third step is repeated till the solution is acceptable for decision maker.

5 Back to the capital market – practical application

Firstly for the multiple criteria evaluation process in each group of shares funds, we set the weights of criteria on the basis of linguistic variables as forenamed above. So the weight of return is random variable with uniform distribution $R(0,6;1)$ and the weight of risk criterion is also random variables with uniform distribution $R(0,8;1)$. Generate 10 weight scenarios, so we get 10 possible rankings of alternatives. The final order is obtained via the assignment problem. The investor chooses only one alternative, consequently in the first place. Selected shares funds are the following - *Dynamický Mix*, *Sporoinvest*, *Bondinvest*, *Global Stocks*.

Secondly we apply the interactive multiple objective programming method proposed above. We have 10 scenarios of aspiration levels of all objective functions. The final mathematical model is formulated as follows

$$\lambda \rightarrow \max$$

$$\begin{aligned} &\sum_{j=1}^{4} v_j x_j - \lambda(U_1 - L_1) \leq L_1 && x_j \geq 0.05 \quad j = 1,...,4 \\ &\sum_{j=1}^{4} r_j x_j + \lambda(U_2 - L_2) \geq U_2 && x_j \leq 0.5 \quad j = 1,...,4 \\ &\sum_{j=1}^{4} n_j x_j + 0.5\lambda \leq 2.5 && \sum_{j=1}^{4} x_j = 1 \\ &\sum_{j=1}^{4} s_j x_j - 0.2\lambda \geq 0 && 0 \leq \lambda \leq 1 \end{aligned},$$

where $v_j, r_j, n_j, s_j\ (j = 1,...,4)$ is return, risk, costs and Sharpe ratio of j^{th} shares fund, $x_j\ (j = 1,...,4)$ represents a share of j^{th} fund in portfolio. The values L_1 and U_1, or L_2 and U_2 are the aspiration levels of particular criterion.

We choose the solution with the biggest objective function value for interactive procedure. Its shape is: 22.6 % Dynamický Mix, 22.4 % Sporoinvest, 50 % Bondinvest, 5 % Global Stocks with 1.28 % return and 2.03 % costs, $\lambda = 0.59$. As the portfolio risk is more important than return, so the investor wishes to make better the value of risk, in the concrete decrease at least under 1.9 % level, on the contrary decrease of return by 0.1 % with the same tolerance is acceptable. Thus the two following constraints must be add into the model

$$\sum_{j=1}^{4} r_j x_j \le 1.9 \qquad 1.28 - \sum_{j=1}^{4} v_j x_j + 0.1\lambda \le 0.2.$$

The next solution is: 48.1 % Dynamický Mix, 41.9 % Sporoinvest, 5 % Bondinvest, 5 % Global Stocks with 1.2 % return and 1.9 % costs, $\lambda = 0.55$. The investor still wants to decrease the risk at the expanse of return, below 1.7 % level with the same acceptable decrease return as in the first case. After the model change (similar as in the previous one) by supplement

$$\sum_{j=1}^{4} r_j x_j \le 1.7 \qquad 1.2 - \sum_{j=1}^{4} v_j x_j + 0.1\lambda \le 0.2,$$

the solution looks: 41.1 % Dynamický Mix, 48.9 % Sporoinvest, 5 % Bondinvest, 5 % Global Stocks with 1.08 % return and 1.7 % costs, $\lambda = 0.5$. The next demand on risk cut-down about 0.2 % is not acceptable because of solution infeasibility. The investor agrees with prior one.

6 Conclusion

The ambition of the article is to introduce the investment decision making where we usually meet many elements of uncertainty.

The stochastic weights of evaluative criteria and some criterial values are taken into account as random variables with some continuous probability distribution leading to scenario analysis in the multiple criteria evaluation process where one of the basic problems of linear programming is applied for making final ranking.

The investor uncertain preferences about further important elements in the process are expressed with the aid of fuzzy sets, or the basic operations in the field of fuzzy logics. The weights can be also expressed by this way, but in this article it is not used because the uncertainty values of weights are just included in the multiple criteria evaluation approach. If all relations are linear, the optimal solution of fuzzy decision making problem becomes accessible by linear programming appliance. It is no doubt, that the extension about some nonlinear components is possible.

Finally the multiple criteria objective method is applied in order to make the "optimal" portfolio with a possibility to change the solution on the basis of the investor fuzzy demands during the process.

Acknowledgements

The research project was supported by Grant No. IGA F4/16 /2011, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Černý, M., Glückaufová, D.: *Vícekriteriální rozhodování za neurčitosti*. Academia, Praha, 1987.
[2] Dlouhý, M., Fábry, J., Kuncová, M., Hladík, T.: *Simulace podnikových procesů*. Computer Press, Brno, 2007.
[3] Fiala, P.: *Modely a metody rozhodování*. Oeconomica, Praha, 2008.
[4] Gordon, J. A., Sharpe, W. F.: *Fundamentals of Investments*. Prentice Hall Inc., New Jersey, 1989.
[5] Gupta, M, Bhattacharjee, D.: Multi Objective Problem in Fuzzy Environment where Resources are Triangular Fuzzy Number. *European Journal of Scientific Research* **46** (2010), 99-106.
[6] Hwang, C. L., Yoon, K.: *Multiple Attribute Decision Making. Methods and Applications*. Springer-Verlag, Berlin, 1981.
[7] Investment company Česká spořitelna, accessible from: *http://www.iscs.cz/*, [cit. 30. 4. 2012].
[8] Jablonský, J.: *Programy pro matematické modelování*. Oeconomica, Praha, 2007.
[9] Klir, G. J., Yuan, B.: *Fuzzy Sets and Fuzzy Logic – Theory and Application*. Prentice Hall PTR, New Jersey, 1995.

Criteria for evaluating the significance of transport infrastructure in precedence analysis

Josef Botlík[1], Milena Botlíková[2]

Abstract. The precedence analysis enables to find new links and relationships in the evaluation of selected variables. The most important finding is the impact of changes in the values on the surroundings. The basis for precedence analysis is the definition of system objects, elements and relationships. The system is then transformed into incidence matrices. Because of examining the changes that have occurred in the selected variables in a certain time interval, we analyze the mutual influence of the relative and absolute changes in the values of the selected element to the neighboring elements. Based on relative and absolute values of these changes, we can determine the increases or decreases directions of these variables in the system. Then we are able to examine antecedents (precedences), consequents (successions) and multiple precedences and successions. The problem in assessing the transport infrastructure is the choice of criteria and their weights to determine the significance of transport infrastructure in the region and the subsequent search of the significance of changes in the time interval. In this paper we deal with changes in road infrastructure, increase or decrease of highways, expressways and I, II., and III. class roads. The search of weights for these transport structures is also presented in the paper.

Keywords: precedence, sukcedence, cities with extended authority, infrastructure.

JEL **Classification**: O18, C65
AMS **Classification**: 65C20

1 The method and tools

In the frame of dissertation and few projects there is the analysis of traffic infrastructure impact on a region development. Precedence analysis was used as a method. Relations of selected quantities between the cities with extended authority (CEAs) in Moravian – Silesian region are being analyzed.

If we are working in the system with relations, it is if there is a mutual effect on each other, we can describe structural and time precedences and subsequences of effects. The matrix for precedence description is commonly called the precedence matrix, for capturing subsequences the succedence matrix. Matrices can be binary, and these are being used in non evaluated graphs. In evaluated graphs is either being used transcription into numerical matrices or into combination of a binary matrix and a numerical vector, describing flows. The special case is the flow matrix, in which a write using values of „1“, „0“ and „-1“ is being used.

In precedence matrix (resp. succedence matrix) is value „1“, if a row element is a direct precedence (resp. subsequence) of a column element. In other cases we write nothing into the matrix (in some cases, we write value „0“, if it is necessary for mathematical processing)[3].

Precedence matrices are used s a tool, allowing effective work with heterogeneous quantities. Matrices serve for identification and record of relations and their orientation in the system. The basis for the analysis is a finding of direct relations between analyzed subjects, then the basic unitary flows of tracked quantities are determined on the basis of uneven progress of these subjects in time. The defining of flows was done by comparison of tracked data of elements with a relation (adjacent CEA) in selected times. Data growths or falls were expressed for the data in percent and by comparison of these quantities the course of particular quantity flow was determined. From the identification of growth and fall between CEA the precedences and succedences are determined.

[1] Silesian University in Opava, School of Business Administration in Karvina, Department of Computer Science, University sq. 1934/3, 733 40 Karvina, botlik@opf.slu.cz,
[2] Silesian University in Opava, School of Business Administration in Karvina, Department of Logistics, University sq. 1934/3, 733 40 Karvina, botlikova@opf.slu.cz,
[3] Own principles and mathematical apparatus are more closely described in [1], [3] , [5] a [7].

1.1 Data basis

Calculations were performed upon the data, available in the databases of Directorate of Roads and Highways and Department of Transportation. For model construction, map materials from the portal of public administration of MSR were used. Data come out from planned changes in infrastructure between 2004 to 2013.

As default standards for weight (rate) of CEA infrastructure change determination was used specification of road network to highways, speed roadways, 1st, 2nd, and 3rd roadways, their increase (or decrease) in a region. It was evaluated, if there is a connection of particular roadway and other parts of infrastructure inside and outside CEA, whether there is a connection to an infrastructure with higher significance rank (European, TEN-T[4], transeuropean networks, corridors, etc.). Next factor was the number and placement of slip roads, new roadways integration into regional and local structures (bypasses, through roads, etc.) and creating of particular infrastructure types. Also the negative traffic impacts of transitive character were assessed. A coefficient of region significance was used for determination of flows (density and quality of infrastructure and connection onto other structures)[5] in case that particular CEAs had not construction at disposal in selected period.

A weight table of infrastructure growth for particular CEAs was created on the basis of selected group of standards. The region significance coefficient (density and quality of infrastructure and connection to other structures) was used in case that relevant CEAs did not have construction in selected time period.

1.2 Precedences and multiple precedences

On the basis of determined weights there was calculated the value determining the growth of traffic infrastructure for every CEA. After the calculation of relative changes between adjacent CEAs, the precedence matrix was created. The matrix has 23 rows and columns, rows and columns consist of a listing of 22 CEAs and their surroundings. In every matrix row was a value „1“ entered in the case, that row CEA precedes column CEA in the meaning of tracked quantity (has lower value of this quantity). By calculation of particular precedence matrix power were multiple precedences found. Multiple precedences show the range of tracked quantity changes in particular CEA onto the surrounding. The existence of precedence of particular length points to the existence of a track in the system. This track length is equal to power of particular precedence matrix. During transition between particular CEAs, there is always a growth of particular quantity.

Upon a binary multiplication[6] there are the existences of particular length precedences determined between particular CEAs. By classic multiplication of matrices we will obtain a frequency of these precedences between particular CEAs. For the analysis were powers calculated to the exponent of 8. Further powers of matrices contained only null values.

2 Weight adjustments of particular traffic roadway types

When determining the weights for particular traffic roadway types and their impact on the surrounding, four main variants were examined.

Variant 1

The variant allowed for spinal connections and an existence of roads of higher significance.
Basic weights of particular changes in CEAs were determined as follows: D = 5, R= 4, I. = 3, II. = 2, III. = 1, influence on the surrounding = 2. Changes in regions were points awarded according to their range and edited according to weights. To prevent casual concurrence in adjacent (MRP), the region significance coefficient was added in 0.2 to 1.5 interval, determined on the basis of CEAs inhabitants, location and size.

Variant 2

In next analysis there were adjusted the weights of slip roads. The aim was to allow for connecting of CEA onto spinal roadway. The variant allows for spinal connections and their connection onto existing infrastructure in given locality (number of slip roads). In this case was the weight of existing slip roads increased from a value „3“ to a value „6“.

[4] TEN-T / Transport infrastructure, available at: http://ec.europa.eu/transport/infrastructure/index_en.htm, online 20.10.2011
[5] In more detail for example in literature [6].
[6] In more detail for example in literature [1]

Variant 3

This variant includes the negative impact of a highway in the case, that it is only a transitive roadway. The weight of highway stretches without connection to traffic infrastructure of a region was decreased to „1“, highway stretches with connection to existing infrastructure have still a value of „6“.

Variant 4

Last variant determines the weights so that it concludes to advantage of regional traffic.

Connection of a highway to existing infrastructure = 6 without slip roads, R and I. = 4, II. And III. = 2.

Upon the growth of particular infrastructure type and particular weights it concluded in 2004-2013 period to quantification of traffic infrastructure growths. Values for tracked 4 variants of weights are in the table 1.

ORP	Bílovec	Bohumín	Bruntál	Český Těšín	Frenštát pod Radhoštěm	Frýdek-Místek	Frýdlant nad Ostravicí	Havířov	Hlučín	Jablunkov	Karviná	Kopřivnice	Kravaře	Krnov	Nový Jičín	Odry	Opava	Orlová	Ostrava	Rýmařov	Třinec	Vítkov
number ORP	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
analysis 1	54	76	0.3	24	0.2	55	0.2	7.2	0.2	14	6.2	35	0.1	9.1	59	24	33	0.4	87	0.1	17	4
analysis 2	60	91	0.3	24	0.2	55	0.2	7.2	0.2	14	6.2	35	0.1	9.1	62	24	33	0.4	102	0.1	17	4
analysis 3	28	59	0.3	24	0.2	55	0.2	7.2	0.2	14	6.2	35	0.1	9.1	46	8	33	0.4	74	0.1	17	4
analysis 4	28	64	0.3	26	0.2	58	0.2	10	0.2	18	8.2	40	0.1	12	50	8	44	0.4	85	0.1	22	4

Table 1 Quantification of traffic infrastructure growths

2.1 Precedence changes for particular weight variants

Table 2 shows the numbers of precedences of available lengths. Numbers are stated upon binary and standard multiplication of matrices. In the left part of the table are changes of precedence number without multiplication between identical predecessors and successors. In this case, at the existence of particular length precedence without multiplication between two CEAs is their existence recorded. In the right part are the changes in multiple precedences. In this case is their number recorded at the existence of particular length precedence between two CEAs.

number of precedences given length								
lenght	1	2	3	4	5	6	7	8
analysis 1	62	114	85	53	32	13	2	0
analysis 2	62	102	82	56	31	12	4	0
analysis 3	62	100	63	32	15	4	1	0
analysis 4	62	108	71	43	26	10	3	1

number of multiple precedences given length								
lenght	1	2	3	4	5	6	7	8
analysis 1	62	146	199	184	100	30	4	0
analysis 2	62	130	161	132	69	24	4	0
analysis 3	62	131	135	83	36	12	2	0
analysis 4	62	139	159	120	68	27	7	1

Table 2 Numbers of precedences of different length

It is clear from the Figure 1 that many changes of traffic infrastructure in chosen time interval manifest mostly to distances of second precedences. Variant 3 with allowing for the negative impact of transitive roadways has the lowest number of precedences at all levels of precedence (at all track lengths) and affects the least surrounding CEAs. Largest impacts on direct surrounding have variants 1 and 4, variant 2 has larger influence (in comparison of all variants) on more distant CEAs. For next analysis there are equations of regression added, formed by 6^{th} order polynoms. These equations are used in further analysis for finding a similarity with other examined quantities.

The Figure 2 compares the numbers of multiple precedences. At this evaluation is the largest impact on surrounding at variant 1, smallest impact, again, at variant 3. It is clear from the graph, that largest numbers have precedences of length 3, at variant 1 and 2 even precedences of length 4. Consequently, between particular CEAs, more distant relations occur in larger number of combinations.

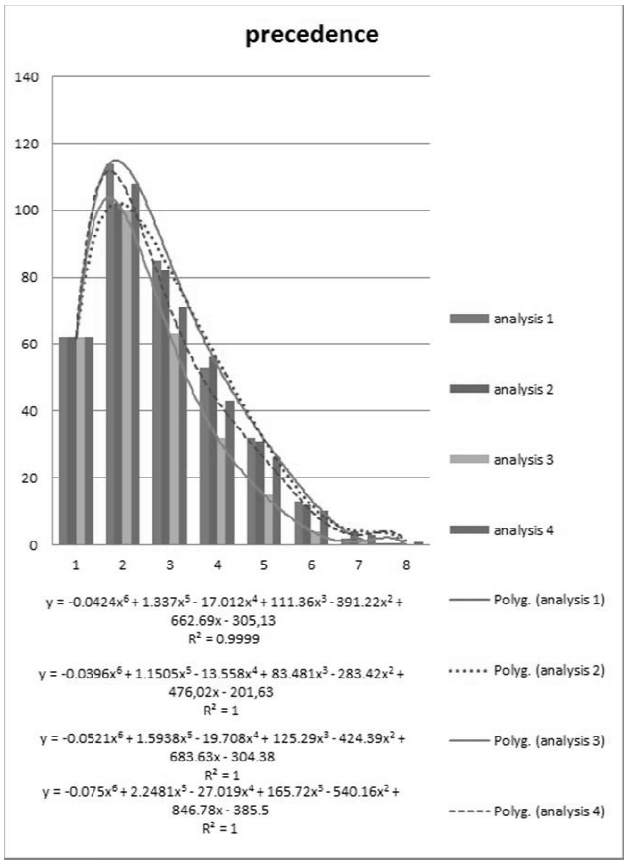


Figure 1 Number of different length precedences

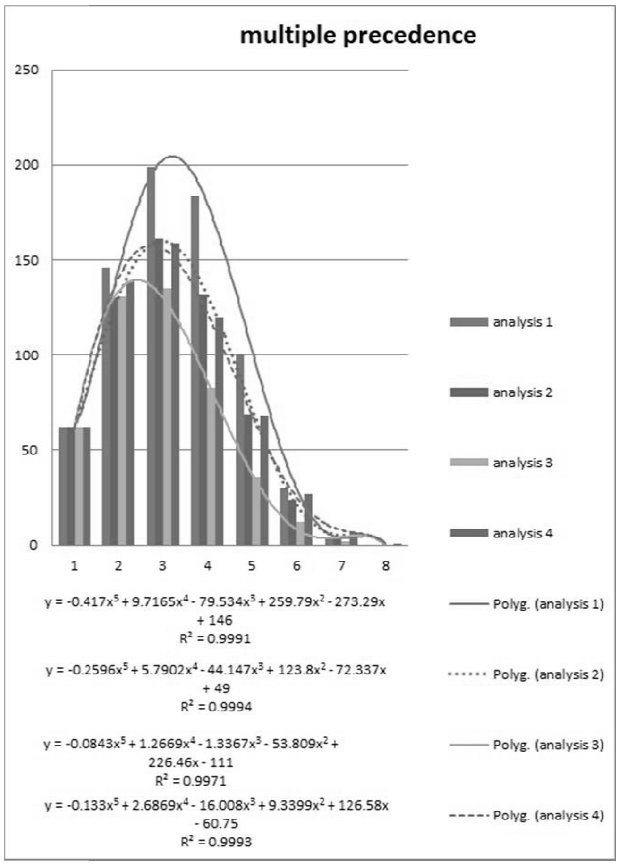


Figure 2 Number of different length multiple precedences

Following graphs represent gradually the changes of impacts according to particular CEAs. These graphs show the number of precedences and multiple precedences between the various CEAs. The first and third line shows the precedences, second and fourth line shows the multiple precedences.

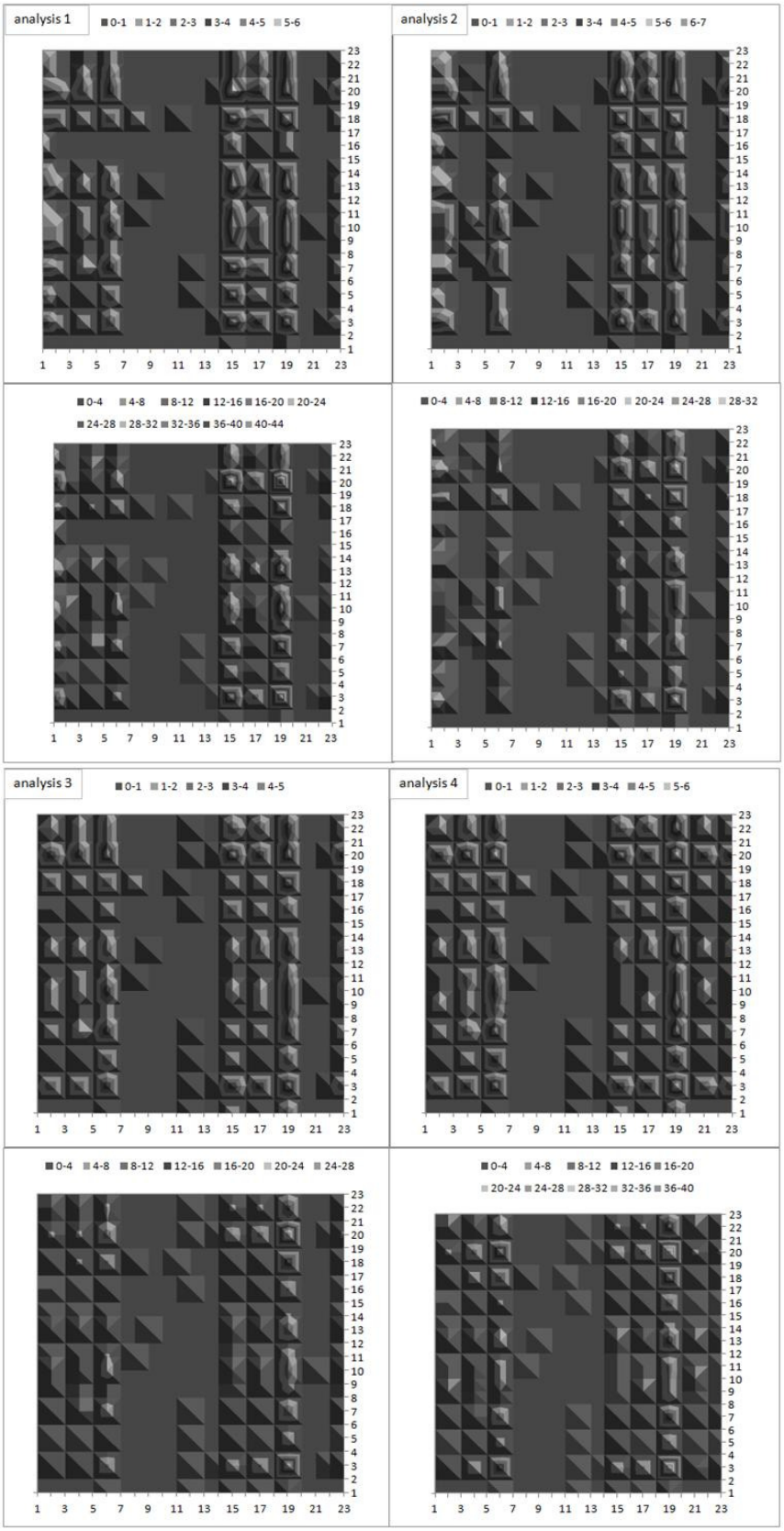


Figure 3 Influence of particular CEAs on different CEAs

3 Conclusion

It is clear from the graphs, that largest impact have the CEAs 1, 2, 6, 12, 15, 17, 19 and it is in all the variants of weights with minimal differences. Largest impact on the surrounding has, therefore, in MSR the construction of D1 and R48 highway.

What is for the whole precedence number, is that the largest impact is recorded at CEAs 6, 15 and 19. It is clear here, that changes of traffic infrastructure are concentrated to large cities.

The proposed method is based examination of change impacts on surrounding, its number and length. The results of analysis are possible to compare with analysis of different, seemingly heterogeneous quantities, at which we examine, again, the number and range of changes. The expression of impacts using the precedences will allow finding similarities in different quantities behavior on surrounding. In relation to investigation of changes in time interval we can analyze also the inertia of changes and time slips between changes at different quantities.

References

[1] Borje, L.: *Theoretical analysis of information systems*. Bratislava: Alfa, 1981.
[2] Botlík, J:. Opportunities of using analytical methods in logistic analysis of region . *Acta Academica* **1** (2011) Karviná: OPF Karviná, 2011, 25-36.
[3] Botlík, J.: Precedence, as a tool for analysis of a system using matrices. In: *Electronic conference E - CON Internactional Masaryk's Conference for doctorands and young scientific workers.* Hradec Králové: MAGNANIMITAS, Hradec Králové, 2010, 292-298.
[4] Botlík, J.: Precedence matrices and their analogy in bipartitionary graphs. In: *PEFnet 2010, 14. class - European Scientific conference of Doctors grade study listeners.* Brno: Mendels University, 2010,
[5] Botlík, J.: Specific analytical tools in logistics. In: *Compilation III. international scientific conference Conference for doctorands and young scientific workers of OPF Karviná.* Karviná, OPF: OPF Karviná, 2010, 65-77.
[6] Botlík, J., Botlíková, M.: Precedental regional analysis in caring economics. In: 7th International Symposium on Business Administration, Canakkale, Turkey, 2012, 206 - 224.
[7] Unčovský, L.: *Modely sieťovej analýzy,* Bratislava: Alfa 1991

The usage of precedence in analysis of impact of the economic crisis for accommodation services

Milena Botlíková [1], Josef Botlík [2]

Abstract. The paper deals with the evaluation of changes in occupancy of accommodation facilities in different regions of the Czech Republic in time of the economic crisis. Input data are obtained from the Czech Statistical Office. In selected time intervals there are detected increases and decreases in values of selected data relating to accommodation services. These changes are compared within each region and are recorded into precedence matrices. Relations captured through precedence are gradually analyzed using the matrix exponentiation. Based on the exponentiation there are identified regions, among which we can observe non decreasing respectively non increasing changes of the monitored values. In more detailed, there are observed precedence of various lengths and found the maximal precedence. The precedence analysis make possible to monitor changes and directions of influence of economic crisis between regions. Comparison of calculated precedence at different time intervals shows the possible dependencies between the variables being monitored in individual regions. The research results can be subsequently used for comparison with changes of other economic variables and the determination of dependencies between these variables.

Keywords: precedence, sukcedence, cities with extended authority, infrastructure.

JEL **Classification**: O18, L83
AMS **Classification**: 65C20

1 Introduction

Paper is based on research projects CZ.1.07/2.3.00/09.0197, SGS/24/2010 and SGS/23/2010. In the projects there were sought generally applicable tools for capturing state variables changes. There were searched easily applicable tools to analyze the dynamic properties of systems.

Basic principles of described method lie in capturing of seemingly heterogeneous quantities and phenomena by their change through time or space, determination of precedences and sequences of these quantities, their description using mathematical apparatus and comparable results generating using standard operations.

A relatively wide opportunity offers the precedence matrix

With precedent and subsequent events, we can analyze the temporal and structural changes in systems. To facilitate the processing is appropriate the use of matrixes in which the link between a line element and a column element is captured. In practice, we use the capture of precedence or succedens of the inline element and the matrices are called precedence (succedens) matrices. Using simple operations on these matrices, we can analyze the existence and frequency of links between elements of the system. The results were published for example in [2], [3], [4] or [5].

2 Model creating

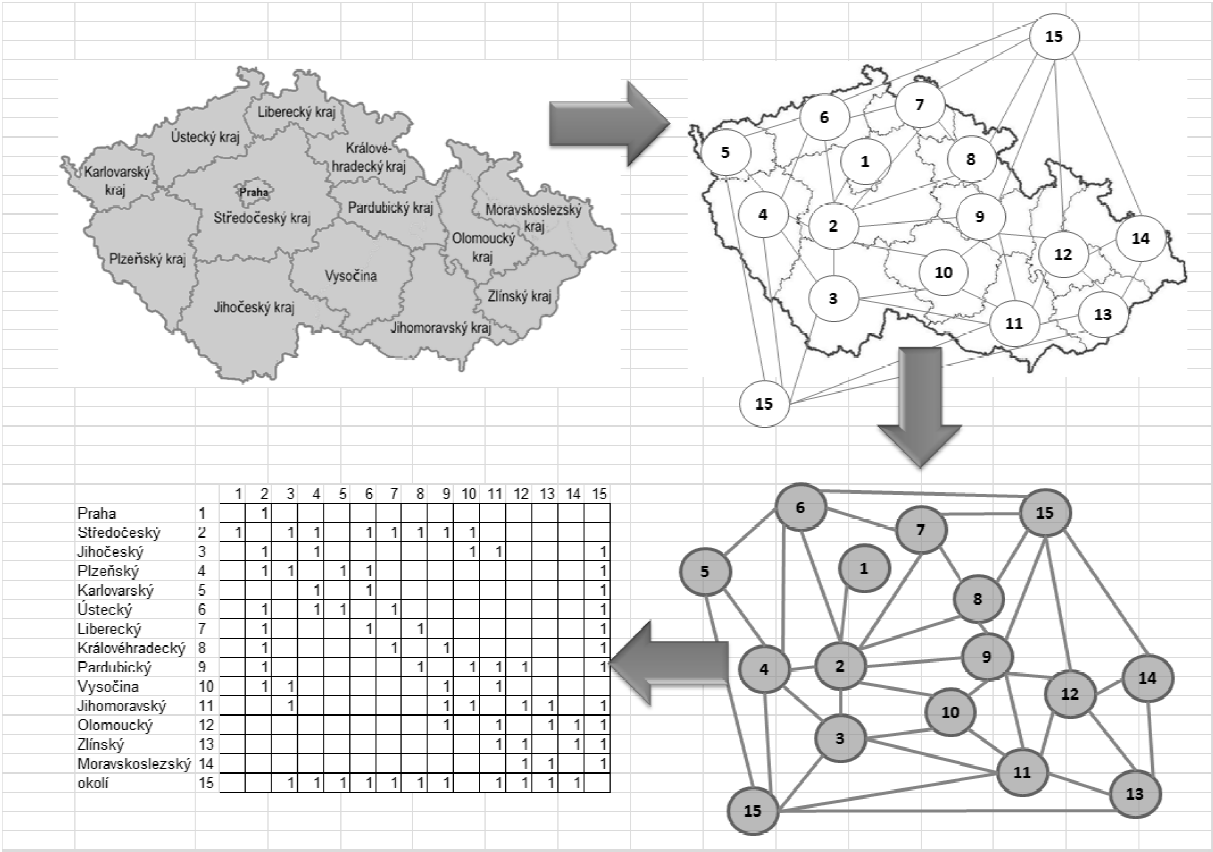
A segmentation of Czech Republic into regions was used for the analysis. On this basis was a simple network graph created, which captures relations between particular regions. Relation is defined by adjacency of particular regions, which have borders with other countries, have defined relation with surrounding.
In this phase of analysis is not captured, with which countries are given regions adjacent with, relation is a common relation to surrounding.

On the basis of network graph is the model transcribed into incidence matrix. In this matrix is the relation structure between particular regions captured. The matrix has defined one row and one column for every region

[1] Silesian University in Opava, School of Business Administration in Karvina, Department of Logistics, University sq. 1934/3, 733 40 Karvina, botlikova@opf.slu.cz
[2] Silesian University in Opava, School of Business Administration in Karvina, Department of Computer Science, University sq. 1934/3, 733 40 Karvina, botlik@opf.slu.cz.

and for surrounding. In every row is a value „1“ in such column, for which there is a relation of region, described by a row and a region described by a column.

As a time interval for the crisis period, was chosen the period from 2008 to 2011 (data for 2012 year are still incomplete). From the data the percentual changes were calculated out in relation to the beginning of interval. Upon comparison of percentual falls and growths in regions, were the courses of particular quantity changes between particular regions determined. Determining of the courses allowed the construction of oriented graphs and creating precedence matrix[3].



		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Praha	1		1													
Středočeský	2	1		1	1		1	1	1	1	1					
Jihočeský	3		1		1						1	1				1
Plzeňský	4		1	1		1	1									1
Karlovarský	5				1		1									1
Ústecký	6		1		1	1		1								1
Liberecký	7		1				1		1							1
Královéhradecký	8		1					1		1						1
Pardubický	9		1						1		1	1	1			1
Vysočina	10		1	1						1		1				
Jihomoravský	11			1						1	1		1	1		1
Olomoucký	12									1		1		1	1	1
Zlínský	13											1	1		1	1
Moravskoslezský	14												1	1		1
okolí	15			1	1	1	1	1	1	1		1	1	1	1	

Figure 1 Creating of incidence matrix

3 Date and analysis

In this phase of analysis there are not the rules for creating of network graph abided by, there are not cycles tracked. Concurrently is the graph considered as an subgraph of higher land structure and abnormalities are not being solved (for example: Prague has a relation only to Mid-Czech region). The values of surrounding are determined, for this analysis, as an average from existing values.

The data from public database of Czech Statistical department was used for this analysis. Upon the growths and falls of guests in accommodation facilities there was analyzed the use of rooms in particular categories of accommodation facilities (hotels “*****”, hotels “****”, hotels “***”, guest houses and other hotels). The following table presents the occupancy of rooms in particular regions and changes in the interval of 2008-2013. From the table is clear lack of data for analysis of “*****” hotels.

[3] More in detail look in literature in references below, for example, in the literature [1] or [6].

	2008	2011	2008-2011	%
1	4 587 483	5 050 923	463 440	10.10
2	665 178	694 083	28 905	4.35
3	936 032	922 237	-13 795	-1.47
4	487 490	496 534	9 044	1.86
5	679 996	703 143	23 147	3.40
6	366 530	336 909	-29 621	-8.08
7	703 787	632 794	-70 993	-10.09
8	902 713	825 621	-77 092	-8.54
9	360 903	323 867	-37 036	-10.26
10	395 820	358 332	-37 488	-9.47
11	1 185 770	1 115 349	-70 421	-5.94
12	426 604	397 279	-29 325	-6.87
13	497 452	469 335	-28 117	-5.65
14	640 128	572 306	-67 822	-10.60
15	916 849	921 337	4 488	0.49

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1		1													
2			1	1		1	1	1	1	1					
3										1	1				
4			1			1									1
5				1		1									1
6							1								
7															
8							1		1						
9															
10									1						
11									1	1		1			
12									1					1	
13											1	1		1	
14															
15			1			1	1	1	1		1	1	1	1	

Figure 2 Creating of precedence matrix (number of guests in accommodation facilities in regions)

No.	regions	hotels *****			hotels ****			hotels ***			other		
		2008	*2011*	*difference*	*2008*	*2011*	*difference*	*2008*	*2011*	*difference*	*2008*	*2011*	*difference*
1	Prague	55.9	59.5	3.6	57.0	60.2	3.2	57.0	59.3	2.4	42.7	43.8	1.1
2	Středočeský	0	0	0.0	36.7	34.9	-1.8	36.1	34.0	-2.1	28.0	20.8	-7.2
3	Jihočeský	0	0	0.0	52.4	45.6	-6.8	32.6	30.6	-2.0	24.9	22.6	-2.4
4	Plzeňský	0	0	0.0	22.3	31.7	9.5	34.8	41.3	6.5	25.0	17.8	-7.2
5	Karlovarský	57.4	55.0	-2.5	66.7	66.6	-0.1	48.5	51.3	2.9	26.0	22.1	-3.9
6	Ustecký	0	0	0.0	37.0	38.7	1.7	29.7	27.7	-2.0	20.1	17.2	-3.0
7	Liberecký	0	0	0.0	42.6	45.5	3.0	32.7	35.4	2.7	26.3	24.0	-2.3
8	Královéhradecký	0	0	0.0	54.7	44.1	-10.6	35.9	37.3	1.3	31.3	26.3	-5.0
9	Pardubický	0	0	0.0	30.9	31.3	0.4	30.8	29.5	-1.3	30.2	23.4	-6.8
10	Vysočina	0	0	0.0	30.8	28.9	-1.9	29.0	27.2	-1.8	25.1	21.8	-3.3
11	Jihomoravský	0	0	0.0	43.1	34.2	-8.9	38.8	33.4	-5.4	28.8	22.4	-6.4
12	Olomoucký	0	0	0.0	45.0	31.3	-13.6	34.1	37.5	3.4	25.6	22.4	-3.2
13	Zlínský	0	0	0.0	57.3	54.6	-2.7	42.0	35.6	-6.4	30.3	25.8	-4.5
14	Moravskoslezský	0	0	0.0	40.4	33.5	-6.8	34.6	31.2	-3.4	29.8	24.6	-5.2
15	surrounding	8.1	8.2	0.1	44.0	41.5	-2.5	36.9	36.5	-0.4	28.2	23.9	-4.2

Table 1 Occupancy of rooms in percent

For particular tracked quantities there were gradually calculated powers of precedence matrices and the number of particular length precedences. Upon the number of precedences between particular regions were determined the ratios of precedences and succedences. These show the ratios between the growth and fall of given quantity for particular regions in comparison with other regions.

As a result, followed comparison of precedences of different lengths at specific quantities. On the Figure 3 are precedence matrices for tracked quantities. Resulting numbers of precedences were compared and show changes, to which it led during the crisis period of 2008. The more precedences given region contains, the higher growth (or lesser fall) of particular quantity values during the crisis period it recorded in relation to surrounding regions.

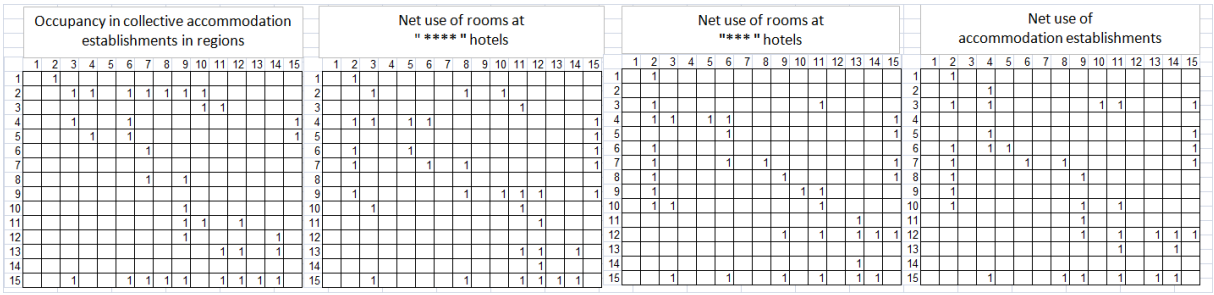
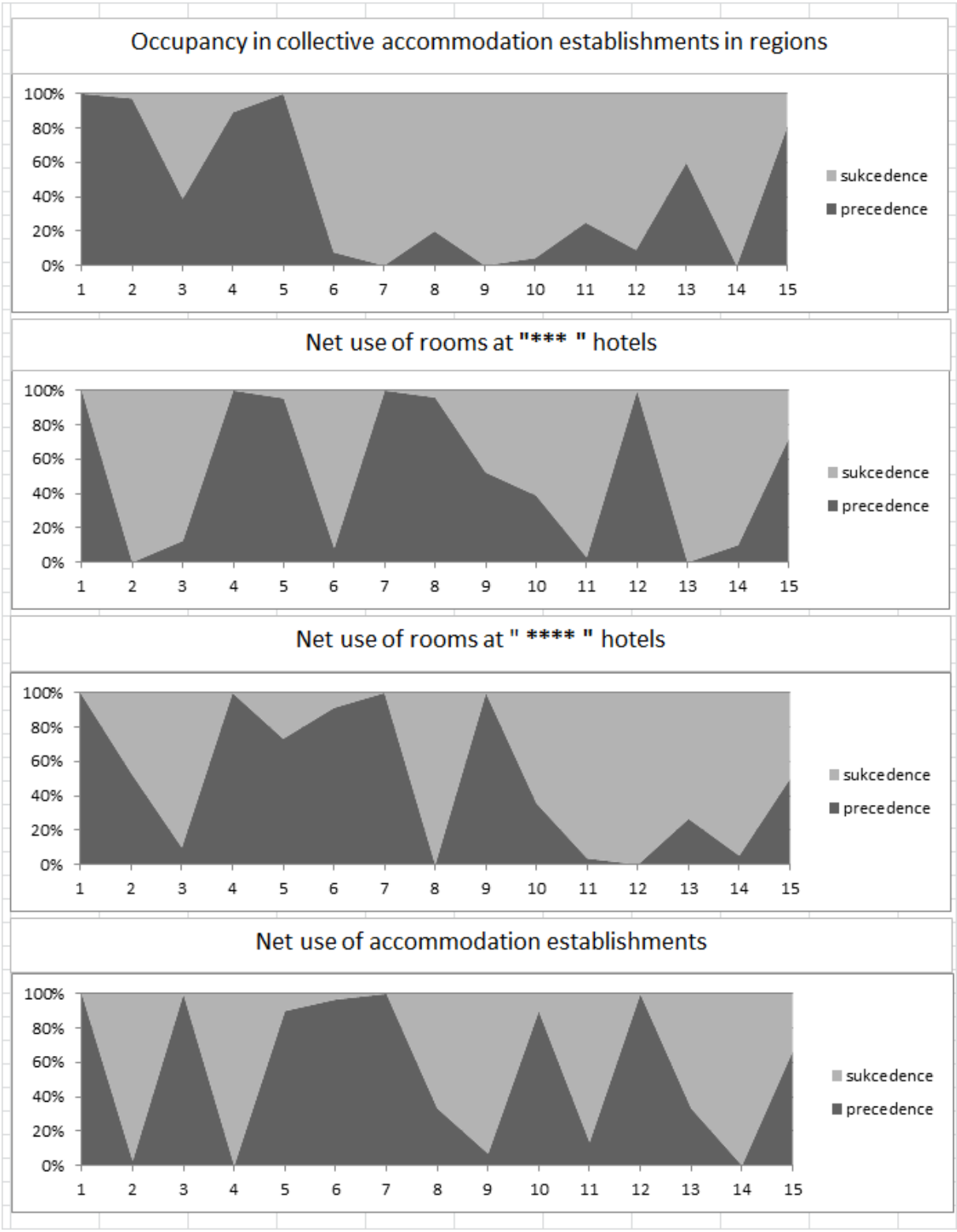


Figure 3 Precedence matrices



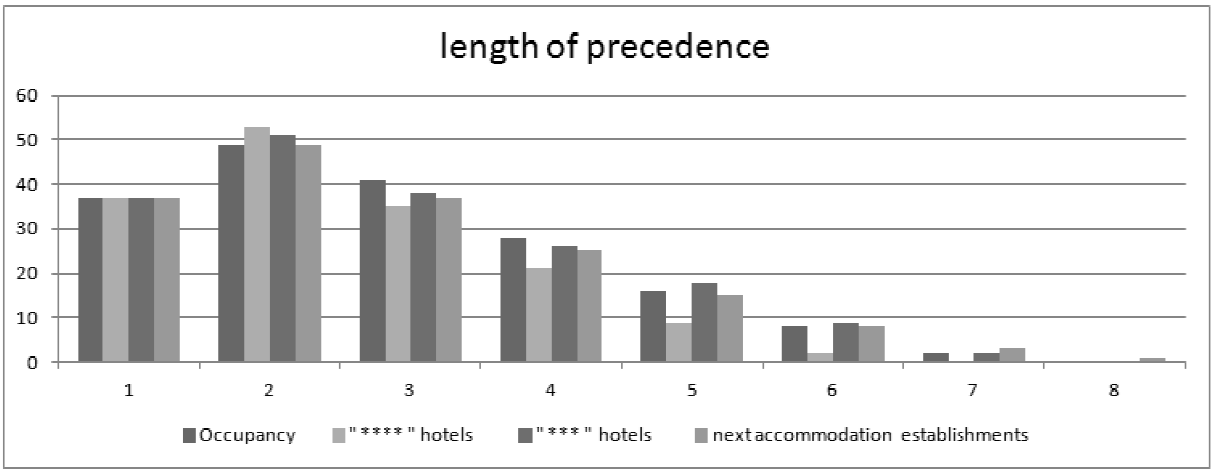
Graph 1 Ratio of precedences and sukcedence for particular regions

It is clear from graph 1, that the largest increase (lowest decrease) of attendance in collective accommodation facilities in regions, recorded Prague, what could be presumed. Relatively large number of precedences is then in Karlsbad region and Pilsen region. The largest fall, in relation to its surrounding, recorded Liberec region, Pardubice region a Moravian – Silesian region. In these regions there is null number of precedences.

Hradec Králové region, South Moravian region, Olomouc region and MSR at the same time recorded a significant fall of room occupancy in higher parameters hotels. Pilsen region has even layout of room occupancy,

in all categories of accommodation facilities it recorded a growth in relation to its surrounding. Interesting is the comparison of Liberec region. This region has larger fall of accommodated, than all the other regions in surrounding, yet it has increase in all types of accommodation facilities. It is clear, that this increase is causes by a larger number of long-term accommodated visitors and smaller capacity of accommodation facilities (even at small number of clients is a high bed use).

Total number of precedences for tracked quantities presents graph 2. On this graph it is visible, that the highest number of precedences is of level 2 precedence, relatively high number of level 3 precedences. It means, that we can follow a relatively lots of tracks between the region of length 2 and 3 transitions, where particular quantities are growing (resp. non-falling). It is clear from the graph, that the largest impact on surrounding regions have the numbers of accommodated in "****" hotels, yet, this impact is on direct surrounding and has not more distant relations. On the other hand, occupancy of beds in *** hotels and other accommodation facilities has the impact less intensive, but of wider range.



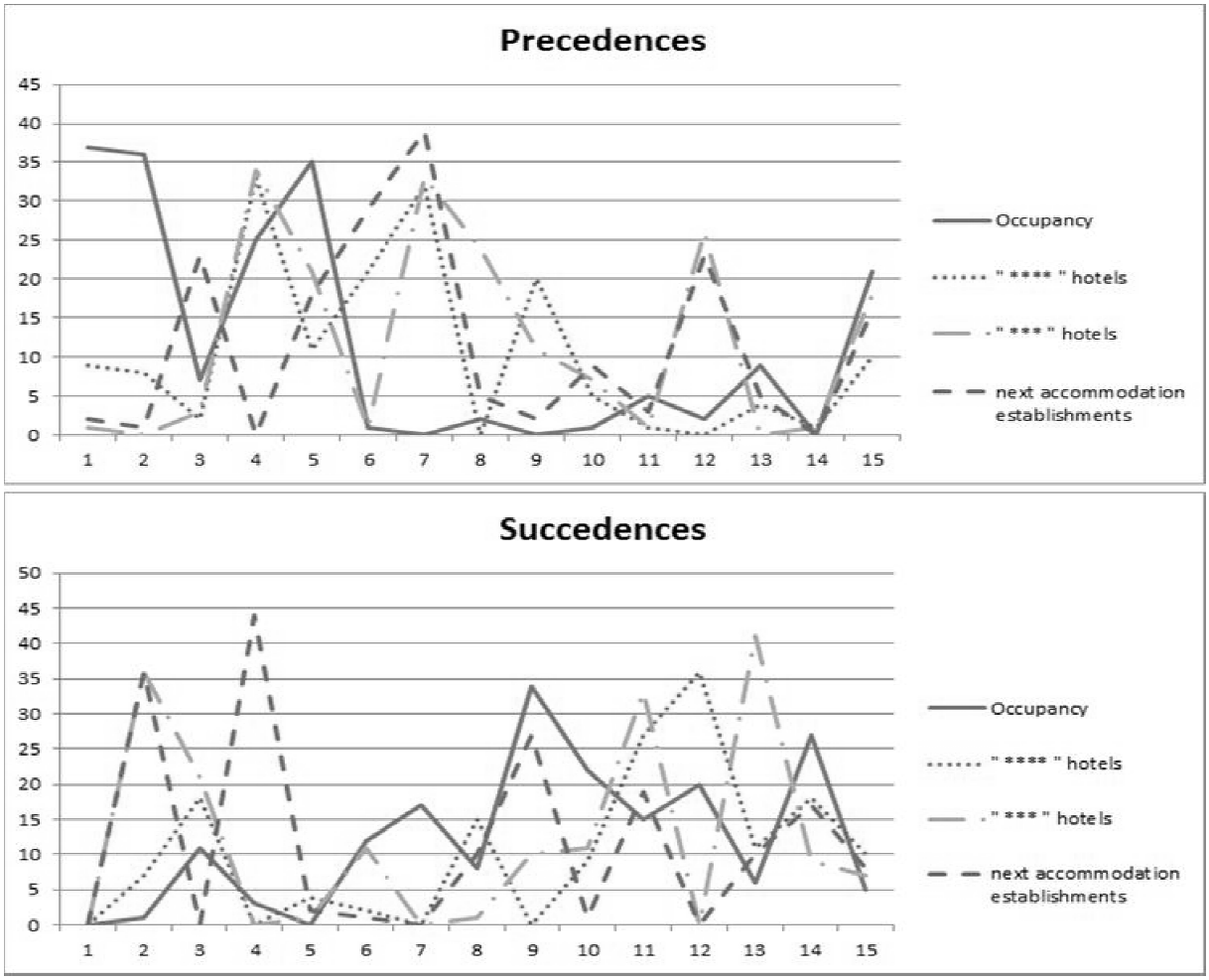
Graph 2 length of precedences

The last table shows the totals of precedences and succedences at tracked quantities in particular regions.

precedences															
region number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Occupancy	37	36	7	25	35	1	0	2	0	1	5	2	9	0	21
" **** " hotels	9	8	2	33	11	21	32	0	20	5	1	0	4	1	10
" *** " hotels	1	0	3	34	21	1	33	24	11	7	1	26	0	1	18
next accommodation establishments	2	1	23	0	18	29	39	5	2	9	3	23	5	0	16
succedences															
region number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Occupancy	0	1	11	3	0	12	17	8	34	22	15	20	6	27	5
" **** " hotels	0	7	18	0	4	2	0	15	0	9	27	36	11	18	10
" *** " hotels	0	36	21	0	1	11	0	1	10	11	33	0	41	9	7
next accommodation establishments	0	36	0	44	2	1	0	10	27	1	19	0	10	17	8

Table 2 Precedences and succedences Czech Republic regions

Graph 3 presents in detail the total of precedences and succedences for particular regions in tracked quantities.



Graph 3 detail of precedences and succedences

4 Conclusion

The contribution shows opportunities of precedence analysis of selected quantities. It is clear, that this way of analysis is able to use at tracking of intensity and range, by which the change of particular quantity influences its surrounding. Intensity is measurable by the number of precedences, the range by the level of precedence. It is suitable to extend the method by determination of muliplicated number of precedences, which we reach by classic multiplication of precedence matrices. In this contribution were used operations, defined in [1], which give at the same length precedences between the same elements the existence of precedences, not the number at all.

References

[1] Borje, L.: *Theoretical analysis of information systems*. Bratislava: Alfa, 1981.
[2] Botlík, J:. Opportunities of using analytical methods in logistic analysis of region . *Acta Academica* **1** (2011) Karviná: OPF Karviná, 2011, 25-36.
[3] Botlík, J.: Precedence, as a tool for analysis of a system using matrices. In: *Electronic conference E - CON Internactional Masaryk's Conference for doctorands and young scientific workers.* Hradec Králové: MAGNANIMITAS, Hradec Králové, 2010, 292-298.
[4] Botlík, J.: Precedence matrices and their analogy in bipartitionary graphs. In: *PEFnet 2010, 14. class - European Scientific conference of Doctors grade study listeners.* Brno: Mendels University, 2010,
[5] Botlík, J., Botlíková, M.: Precedental regional analysis in caring economics. In: 7th International Symposium on Business Administration, Canakkale, Turkey, 2012, 206 - 224.
[6] Unčovský, L.: *Modely sieťovej analýzy,* Bratislava: Alfa 1991

Analysis of factors influencing toll amount collected on the Czech roads

Milena Botlíková,[1] Šárka Čemerková[2],

Abstract. The problem of fee for using the highways and first class roads (toll), collected in the Czech Republic is investigated in this paper. Legal basis for the tolling are the "Act on the toll", "notice on the using charging roads", and also in a near future forthcoming "Act to extend the toll". Currently State Fund of Transport Infrastructure (SFTI) faces a lack of funding; therefore it is necessary to analyze the various items of income for subsequent determination of the amount. Tolls as one of the most important component part of the revenue budget SFTI of the Czech Republic determine the further development of transport infrastructure. The development must be in accordance with Transport Policy of the Czech Republic and European agreements. Collected toll amount may be influenced by different factors e.g. the number of registered vehicles, road transport performance, or intensity of traffic, etc. Because the influence of these factors on the toll varies, it is necessary to determine the significance and the relationships between these factors. Analysis of the toll impact can then be used to determine the correct amount of tolls. This will create proper conditions for further development of sustainable transport.

Keywords: toll, road, analysis, influence, infrastructure, transport.

JEL Classification: C44, C01, C51
AMS Classification: 62P20, 62J12, 91B70

1 Introduction

Development of transport infrastructure is earmarked as one of the priorities of the strategic plans of the CZ. According to many sources indicated the development of road infrastructure as one of the most important instruments having a significant multiplier effect leading to the stimulation of economic development of the CZ. Construction and modernization of road infrastructure experienced a big boom in recent years. The construction of many roads of national and European nature has happened during the past years. The need to build high-quality transport infrastructure is resulting from a strategic position of the CZ and the need for the road network connecting European transport corridors.

Original intentions have changed in recent years. Due to political instability and economic situation, there was a reason they are impeding the development of decreasing financial resources. State Transport Infrastructure Fund, which provides financing and development of transport infrastructure, finance and disposing of state funds from the European Union, recorded since 2008, is decreasing national expenditure framework. Insufficiency cover construction and modernization of transport infrastructure and national resources in recent years has been dealt with by EU finance and the CZ exceeded the principle of additionality, which would ensure that funds will not be from the European Structural Funds used to replace national structural subsidies for public financing and similar expenditure would otherwise be funded from the state budget.

Although the convergence programs of governments define the development and modernization of transport infrastructure as a priority to increase the competitiveness of the CZ in the EU, with the advent Nečas governments seeking to reduce the public deficit has been a drastic reduction of funds for development and much of the transport structures remained in various stages under construction, some projects were not started at all. For example, in 2010, was stopped 12 vehicles, including the strategically important crossroads Silesian cross forming in the European transport corridor.

.

[1]Silesian University in Opava, School of Business Administration in Karvina, department of Logistics, Univerzitní nám. 1934/3, 733 40 Karviná, Czech Republic, botlikova@opf.slu.cz .
[2] Silesian University in Opava, School of Business Administration in Karvina, department of Logistics, Univerzitní nám. 1934/3, 733 40 Karviná, Czech Republic, cemerkova@opf.slu.cz.

2 The toll

Result of efforts to increase national expenditure framework has been gradually in previous years to increase the income of individual items of the State Transport Infrastructure Fund. Over the years, payments from income were gradually increasing, i.e. income from fuels and lubricants, the signs and highway tolls.

The introduction of tolls in 2007, should contribute not only to obtain funds to fund transport infrastructure [5], but also to fulfill the control function, to eliminate congestion and externalities in the traffic flows, when the introduction of a toll in Germany and Austria became the Czech Republic – one of the transit countries.

Toll rates and their changes

Pay toll payment for actual mileage, the amount of which is differentiated according to the season and the environmental burden, and in recent years has seen several changes. Individual changes in the above toll from both those related to highways, and on the expressways and roads. In completing individual sections of the motorway and road network is to expand the number of sections subject to toll charges. Since 1997, when there were about 1 160 km charged roads, it is now charged about 1300 km road network (which is about 2% of the total length of road infrastructure). In 2010 the extension of the payment obligations to vehicles from 3.5 tons was realized. Until this year there was obligation to pay a toll only on vehicles from 12 tones. Other changes in toll scope can be seen in Table 1.

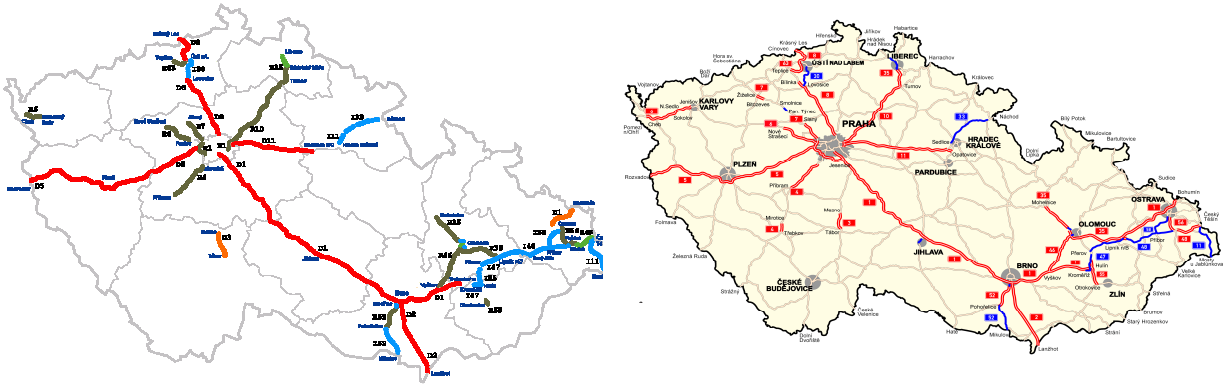
2011	2012
Increase in toll rates by 25%	Increase in toll rates by 25%
Implementation of a compensation system	Possible introduction of 20% VAT
Online system for monitoring of arrears	Quantity discounts
Friday discount on toll rates	

Table 1 Current changes in toll

Between 2010 and 2011, there were no changes in the amount of toll. The increase is related to that year only runs Friday, which was connected with an effort to reduce freight during weekend days. Years 2012 has seen an increase not only in Friday, but even ordinary days and were compared to previous years vehicles to charge EURO V and above, which by this time were exempt from tolls. Toll rates are shown in the following table 2.

		highway			first class roads		
Emissions	**Axle vehicle**	**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
			Ordinary/Friday	Ordinary/ Friday		Ordinary/ Friday	Ordinary/ Friday
EURO II	**2**	2.3	2.26 / 2.88	3.34 / 4.24	1.1	1.08 / 1.37	1.58 / 2.00
	3	3.7	3.63 / 5.55	5.67 / 8.10	1.8	1.77 / 2.70	2.74 / 3.92
	4 +	5.4	5.30 / 8.10	8.24 / 11.76	2.6	2.55 / 3.9	3.92 / 5.60
EURO III and more	**2**	1.7	1.7 / 2.12	2.61 / 3.31	0.8	0.79 / 1.00	1.23 / 1.56
	3	2.9	2.9 / 4.35	4.45 / 6.35	1.4	1.37 / 2.10	2.14 / 3.06
	4 +	4.2	4.2 / 6.30	6.44 / 9.19	2	1.96 / 3.00	3.06 / 4.38
EURO V+	**2**			1.67 / 2.12			0.79 / 1.00
	3			2.85 / 4.06			1.37 / 1.96
	4			4.12 / 5.88			1.96 / 2.8

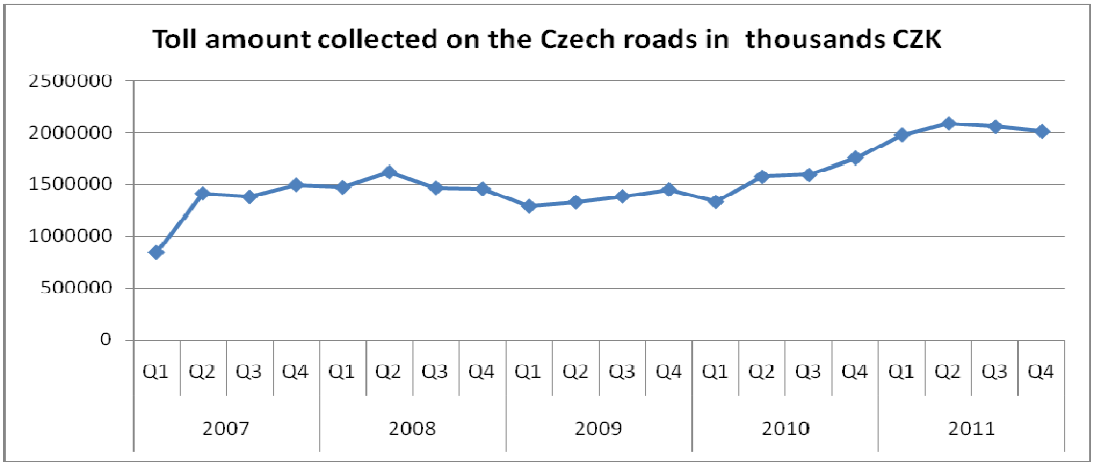
Table 2 Rates of toll (in CZK/km) [8]



Picture 1 Czech sections Toll roads in 2008 and 2012 [6]

Increase in toll rates along with the increase in the number of kilometers of toll, however, did not bring the desired effect. The increase in revenues of the State Fund for Transport Infrastructure Development of the toll is still not sufficient and in the second half of 2011 declined.

Many analyses are based on the assertion that if the toll increases further, the individual carriers, burdening the already high price of fuel, try to search cost-free route – route without tolls. Of the factors that need further analyze the different factors that selected toll on the amount of influence.



Picture 2 Toll amount collected on the Czech roads, in thousands CZK [4]

3 Factors affecting the selected toll and finding a mathematical model

In search of factors affecting the amount of tolls the following factors were selected: price of diesel at filling stations in the CR (X_1 explanatory variable), transport performance (variable X_2) and the number of goods vehicles of categories N2 and N3 split into second-hand vehicles (X_3) and the newly registered vehicles (X_4). These factors were selected for the following reasons:

- The price of oil affects the volume of traffic.
- It is assumed that increasing the transport performance will match the growing volume of selected toll.
- The old trucks do not meet the higher standards and meet the EURO higher toll rates. It can be assumed that with the increasing number of these vehicles there will be selected higher toll.
- In contrast, newly registered vehicles are mostly new vehicles meet the highest standard EURO and they therefore correspond to the lowest toll rate, i.e. with the increasing number of these vehicles at the expense of used cars it will reduce the toll.

Since some factors are only available quarterly data for the other factors were adjusted so that the data corresponded to the same time sections. From the daily price of oil released by the Czech Statistical Office [4]were calculated quarterly arithmetic averages and were calculated for the other quarterly totals.

The aim was to find multiple linear regression model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 . \tag{1}$$

The advantage of linear economic models is particularly illustrative in the economic interpretation and the possibility to estimate and test the standard and simple procedures, which cannot always tell the models that are linear in the parameters.

First, a linear relationship has been studied above the selected toll on oil prices. Found regression function has the form (model 2)

$$Y = -768940.31 + 77184.9 X_1 \tag{2}$$

with the value of the coefficient of determination 0.636, adjusting coefficient of determination 0.616, *p*-value of the individual *t*-test of parameter β_1 is $2.5 \cdot 10^{-5}$.

The final model was then tested. In case accaptance of the model with criterion „price of fuel“ then the model (2) is autocorelated (d=1.38).

$$DW = d = \frac{\sum_{t=2}^{n} \left(e_t - e_{t-1}\right)^2}{\sum_{t=1}^{n} e^2} \tag{3}$$

where: e…..residue

Other assumption for verification of significance of the regressive model is satisfying homoscedasticity condition. This test for model (2) was done using Golfeld – Quandt test (C-Q test [1] by comparing test criterion (4) and critical value (5).

$$F = \frac{s_1^2}{s_2^2} \Rightarrow s_1^2 \rangle s_2^2 \tag{4}$$

where: F …..test criterion

s^2…..variance

$$F_{1-\frac{\alpha}{2}}\left(q-p, q-p\right) \tag{5}$$

where: q – p.. *when* q – number of observations, p – number of parameters

α …level of significance

Homoscedasticity is confirmed when:

$$F \rangle F_{1-\frac{\alpha}{2}} \tag{6}$$

Using G-Q test (6) there was confirmed homoscedasticity (4.05 < 5.05)

To increase the relevancy of the model the second explanatory variable (vehicle performance) was added. Found regression function has the form (7)

$$Y = -852115.83 + 46988.62 X_1 + 79.95 X_2 . \tag{7}$$

The inclusion of this explanatory variable in the model has only a slight increase in the coefficient of determination (0.698). Since the *p*-value of the individual *t*-test for parameter β_2 was 0.079 and the for variables X_1 and X_2 was found a high correlation coefficient value (0.782, see Table 3), second explanatory variable was excluded from the model due to multicollinearity.

	Transport performace X_2 (in tshousands.tonnes)	Average price of diesel in CZK (X_1)
Transport performaceX_2 (in tshousands.tonnes)	**1**	
Average price of diesel in CZK (X_1)	**0.781868**	**1**

Table 3 Correlation coefficient value

If we consider the above-selected tolls dependence on oil prices standing and registered second-hand trucks, we obtain the regression function of the form

$$Y = -371218.72 + 80823.83X_1 - 620.19X_3 . \quad (8)$$

Other results are shown in the following table 4.

correlation coefficient R	0.88	p-value F test	$3.16*10^{-6}$
coefficient of determination R^2	0.775	adjusting coefficient of determination $adjR^2$	0.748
p-value for β_1	$1.44*10^{-6}$	p-value for β_3	0.005

Table 4 Quality of the regression function

The fourth explanatory variable X_4, newly registered vehicles, was added to the model last. Found regression function has the form

$$Y = -181539.01 + 76038.06X_1 - 857.86X_3 + 73.47X_4 . \quad (9)$$

Delivery of the explanatory variables in the model did not increase the coefficient of determination ($R^2 = 0.783$, $adjR^2 = 0.742$). The p-value of the individual t-test for the parameter β_4 is 0.431. Because between second and newly registered trucks there was also found a correlation coefficient 0.813 (see Table 5), the fourth variable was also excluded from the model.

	Newly registered vehicles (X_4)	Second-hand vehicles (X_3)
Newly registered vehicles (X_4)	**1**	
Second-hand vehicles (X_3)	**0.813362**	**1**

Table 5 Correlation coefficient value

From the originally planned four explanatory variables only variables X_1 and X_3 were finally included in the model and the resulting regression function has the form (8).

This regression function was then examined in terms of heteroscedasticity using the Spearman correlation test sequence, using the Durbin-Watson test (3, 11) was examined first order autocorrelation.

$$DW = d = \frac{\sum_{t=2}^{n} \left(809\,216.391\right)^2}{\sum_{t=1}^{n} 631675.713^2} = 1.6411 \quad (11)$$

In both cases, the function appears to be satisfactory. The perimeter of the regression function is also well suited in multicollinearity (see Table 6).

	Second-hand vehicles (X_3)	Average price of diesel in CZK (X_1)
Second-hand vehicles (X_3)	**1**	
Average price of diesel in CZK (X_1)	**0.100556**	**1**

Table 6 Correlation coefficient value

4 Conclusion

After examining the various factors depending on the toll based on multiple linear regression, which allows easy interpretation of results was found to be above the carrying capacity of newly registered vehicles and shall not affect the tolling. On the contrary, the price of oil and the amount of used commercial vehicles related to its size. According to the resulting model it is thus clear that if the unit price will rise the price of fuel (diesel) then the amount of the national budget framework will grow to 80 823.83 CZK as well as an increase of unit used vehicles then decreases the amount of the toll – 620.19 CZK.

The positive effect is probably caused by trying to minimize transportation costs. Using routes is paid not only to shorten transport routes but also reduce transport times, which also allows transport operators to increase competitiveness. Whether the positive effect of oil price increases in the total toll is to be determined by assessing the mutual impact of toll levying excise taxes on fuel. But it is not clear whether increased toll on the fuel price increase will cover any drop in income tax on diesel fuel caused by reluctance to buy at a higher price.

In the case of reducing the toll and increasing the number of used vehicles, it can be concluded that these carriers do not use paid sections. It is obvious that these vehicles trying to use the roads leading cities and communities and thus negatively contribute to environmental quality.

References

[1] Čechura, L. a kol.: *Cvičení z ekonometrie*. ČZU, Praha, 2009, 2nd edition, 102.
[2] Hušek,R.: *Ekonometrická analýza.* VŠE, Praha, 2007, 1st edition, 368.
[3] Hušek, R., Pelikán, J.: *Aplikovaná ekonometrie: teorie a praxe*. Professional Publishing, Praha, 2003, 1st edition, 263.
[4] Ministry of Transport, http://www.mdcr.cz/cs/Statistika_dopravy/default.htm, online 2012.
[5] Ministry of Transport, http://www.mdcr.cz/cs/Silnicni_doprava/Silnice+dalnice+mosty/mytne/
[6] State Fund of Transport Infrastructure, http://www.sfdi.cz/CZ/2012dk-zpoplatnene-useky.php, online 2012
[7] The association of motor carriers, http://cesmad.tradecentrum.cz/, online 2012.
[8] Toll Rates in 2012, http://www.businessinfo.cz/cz/clanek/doprava/mdcr-sazby-mytneho-v-roce-2012/1000470/62871/, online 2012.
[9] Tvrdoň, J. *Ekonometrie.* ČZU, Praha, 2008, 5th edition, 228.

Underwriting risk control in non-life insurance via generalized linear models and stochastic programming

Martin Branda [1]

Abstract. We focus on rating of non-life insurance contracts. We employ multiplicative models with basic premium levels and specific surcharge coefficients for various levels of selected segmentation criteria (rating factors). We use generalized linear models to describe the probability distribution of total losses for a contract during one year. In particular, overdispersed Poisson regression is used to model the expected number of claims during a given period and Gamma or Inverse-Gaussian regression are applied to predict average claim severity. We propose stochastic programming problems with reliability type constraints for the surcharge coefficients estimation which take into account riskiness of each rate cell, prescribed loss ratio and other business requirements. We apply the approach to Motor Third Party Liability (MTPL) policies.

Keywords: non-life insurance, rate making, generalized linear models, stochastic programming, MTPL.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Traditional credibility models in non-life insurance take into account known history of a policyholder and project it into policy rate, see [8]. However, for new business, i.e. new clients coming for an insurance policy, the history need not to be known or the information may not be reliable. Thus traditional approaches to credibility can not be used. We will employ models which are based on settled claims of new contracts from the previous years. This experience is transfered using generalized linear models (GLM), see [14], which cover many important regression models and are nowadays widely applied in insurance, cf. [1, 9, 12, 15]. Expected claim count on a policy during one year and expected claim size can be explained by various independent variables which can serve as segmentation criteria, e.g. age and gender of the policyholder and region where he or she lives. Using these criteria and GLM we can derive surcharges which enable to take into account riskiness of each policy. However, as we will show in this paper, these coefficients need not to fulfill business requirements, for example restriction on maximal surcharge. Optimization models must be then employed.

Stochastic programming techniques can be used to solve optimization problems where random coefficients appear. In this paper, we will employ formulation based on reliability type constraints such as chance constraints and the reformulation based on one-sided Chebyshev inequality. The distribution of the random parts will be represented by compound Gamma-Poisson and Inverse Gaussian-Poisson distributions with parameter estimates based on generalized linear models. Sensitivity analysis of the results with respect to the underlying distribution is often proposed, cf. [2, 6, 7, 10, 13].

This paper is organized as follows. In Section 2, we will review definition and basic properties of generalized linear models. Rate-making approach based on GLM is then proposed in Section 3. In Section 4, optimization models for rates estimation are introduced which enable to take into account various business requirements. These models are extended using stochastic programming techniques in Section 5. Section 6 concludes the paper with an application of the proposed methods to MTPL contracts.

[1]Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Praha 8, branda@karlin.mff.cuni.cz

2 Generalized linear models

In this section, we introduce generalized linear models (GLM) which cover many useful regression models. GLM are based on the following three building blocks:
1. The dependent variable Y_i has distribution from the exponential family with probability density function

$$f(y;\theta_i,\varphi) = \exp\left\{\frac{y\theta_i - b(\theta_i)}{\varphi} + c(y,\varphi)\right\}, \tag{1}$$

where b, c are known functions and θ_i, φ are unknown canonical and dispersion parameters.
2. A linear combination of independent variables is considered

$$\eta_i = \sum_j X_{ij}\beta_j, \tag{2}$$

where β_j are unknown parameters and X_{ij} are given values of predictors.
3. The dependency is described by a link function g which is strictly monotonous and twice differentiable

$$\mathbb{E}[Y_i] = \mu_i = g^{-1}(\eta_i). \tag{3}$$

The most important members of the exponential family are proposed in Table 1 including basic characteristics which are introduced below. The following relations can be obtained for expectation and variance under the assumption that b is twice continuously differentiable

$$\mathbb{E}[Y] = b'(\theta), \tag{4}$$
$$var(Y) = \varphi b''(\theta) = \varphi V(\mu), \tag{5}$$

where the last expression is rewritten using the variance function which is defined as $V(\mu) = b''[(b')^{-1}(\mu)]$.

Distribution	**Density** $f(y;\theta,\varphi)$	**Dispersion param.** φ	**Canonical param.** $\theta(\mu)$	**Mean value** $\mu(\theta)$	**Variance function** $V(\mu)$
$N(\mu,\sigma^2)$	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(y-\mu)^2}{2\sigma^2}}$	σ^2	μ	θ	1
$Po(\mu)$	$\frac{\mu^y e^{-\mu}}{y!}$	1	$\log(\mu)$	e^θ	μ
$\Gamma(\mu,\nu)$	$\frac{1}{\Gamma(\nu)y}\left(\frac{y\nu}{\mu}\right)^\nu e^{-\frac{y\nu}{\mu}}$	$\frac{1}{\nu}$	$-\frac{1}{\mu}$	$-\frac{1}{\theta}$	μ^2
$IG(\mu,\lambda)$	$\sqrt{\frac{\lambda}{2\pi y^3}}e^{-\frac{\lambda(y-\mu)^2}{2\mu^2 y}}$	$\frac{1}{\lambda}$	$-\frac{1}{2\mu^2}$	$\frac{1}{\sqrt{-2\theta}}$	μ^3

Table 1 Distributions from the exponential family

Maximum likelihood method is used to estimate the parameters of GLM. For overdispersed Poisson model where the variance need not to be equal to the expected value (dispersion φ is not set to 1 but is estimated from data) quasi-likelihood function must be used. For details see [14].

3 Rate-making using generalized linear models

We denote $i_0 \in \mathcal{I}_0$ the basic segmentation criterion, e.g. tariff groups, and $i_1 \in \mathcal{I}_1, \ldots, i_S \in \mathcal{I}_S$ the other segmentation criteria which should help us to take into account underwriting risk. We will denote one risk cell $I = (i_0, i_1, \ldots, i_S)$ with $I \in \mathcal{I} = \mathcal{I}_0 \otimes \mathcal{I}_1 \otimes \cdots \otimes \mathcal{I}_S$. Let $L_I = \sum_{n=1}^{N_I} X_{In}$ denote aggregated losses over one year for risk cell I where N_I is the random number of claims and X_{In} the random claim severity. All the variable are assumed to be independent. Then, for the mean and the variance it holds

$$\mu_I = \mathbb{E}[L_I] = \mathbb{E}[N_I]\mathbb{E}[X_I], \tag{6}$$
$$\sigma_I^2 = var(L_I) = \mathbb{E}[N_I]var(X_I) + (\mathbb{E}[X_I])^2 var(N_I). \tag{7}$$

The premium is based on multiplicative model composed from basic premium levels Pr_{i_0} and non-negative surcharge coefficients (rating factors) $e_{i_1}, \ldots, e_{i_S}$, i.e.

$$Pr_I = Pr_{i_0} \cdot (1 + e_{i_1}) \cdot \cdots \cdot (1 + e_{i_S}). \tag{8}$$

Our goal is to find the optimal basic premium levels and coefficient with respect to a prescribed loss ratio $\hat{LR}$, i.e. we would like to fulfill the constraints

$$\frac{L_I}{Pr_I} \leq \hat{LR} \text{ for all } I \in \mathcal{I}. \tag{9}$$

The goal loss ratio $\hat{LR}$ is usually based on management decision. It is possible to prescribe different loss ratios for each tariff cell but this is not considered in this paper. However, losses L_I are random. The simplest way is to hedge against expected value of losses $\mathbb{E}[L_I]$. This can be done directly using GLM with logarithmic link function.

Poisson distribution and Gamma or Inverse Gaussian without intercept are used to estimate parameters for expected number of claims and severity. If we use logarithmic link function in both regression models, then we can get for $I = (i_0, i_1, \ldots, i_S)$

$$\mathbb{E}[N_I] = \exp\{\lambda_{i_0} + \lambda_{i_1} + \cdots + \lambda_{i_S}\}, \tag{10}$$
$$\mathbb{E}[X_I] = \exp\{\gamma_{i_0} + \gamma_{i_1} + \cdots + \gamma_{i_S}\}, \tag{11}$$

where λ_i, γ_i are estimated coefficients. Thus for the mean loss it holds

$$\mathbb{E}[L_I] = \exp\{\lambda_{i_0} + \gamma_{i_0} + \lambda_{i_1} + \gamma_{i_1} + \cdots + \lambda_{i_S} + \gamma_{i_S}\}. \tag{12}$$

Now, if we set $\hat{\lambda}_i = \exp(\lambda_i)/min_{i\in I}\exp(\lambda_i)$ and $\hat{\gamma}_i = \exp(\gamma_i)/min_{i\in I}\exp(\gamma_i)$, the basic premium levels and surcharge coefficient can be estimated as

$$Pr_{i_0} = \frac{\exp\{\lambda_{i_0} + \gamma_{i_0}\}}{\hat{LR}} \prod_{s=1}^{S} \min_{i\in\mathcal{I}_s} \exp(\lambda_i) \prod_{s=1}^{S} \min_{i\in\mathcal{I}_s} \exp(\gamma_i), \tag{13}$$
$$1 + e_{i_s} = \exp\{\lambda_{i_s} + \gamma_{i_s}\}, \tag{14}$$

Then the constraints (9) are fulfilled in expectation. However, the surcharge coefficient estimates often violate business requirements, especially they can be too high, as we will show in the numerical study.

4 Optimization problem for rate estimation

The constraints (9) with expectation can be rewritten as

$$\mathbb{E}[L_{i_0,i_1,\ldots,i_S}] \leq \hat{LR} \cdot Pr_{i_0} \cdot (1 + e_{i_1}) \cdot \cdots \cdot (1 + e_{i_S}). \tag{15}$$

There can be set business limitation that the highest aggregated risk surcharge is lower than a given level r^{max}. We would to minimize basic premium levels and surcharges which are necessary to fulfill the prescribed loss ratio and the business requirements. This leads to the following nonlinear optimization problem

$$\begin{aligned} \min \prod_{i_0\in\mathcal{I}_0} Pr_{i_0} \prod_{i_1\in\mathcal{I}_1} (1 + e_{i_1}) \cdot \cdots \prod_{i_S\in\mathcal{I}_S} (1 + e_{i_S}) & \\ \text{s.t.} & \\ \hat{LR} \cdot Pr_{i_0} \cdot (1 + e_{i_1}) \cdot \cdots \cdot (1 + e_{i_S}) &\geq \mathbb{E}[L_{i_0,i_1,\ldots,i_S}],\ (i_0, i_1, \ldots, i_S) \in \mathcal{I}, \\ (1 + e_{i_1}) \cdot \cdots \cdot (1 + e_{i_S}) &\leq 1 + r^{max}, \\ e_{i_1}, \ldots, e_{i_S} &\geq 0. \end{aligned} \tag{16}$$

Using logarithmic transform of the decision variables $u_{i_0} = \ln(Pr_{i_0})$ and $u_{i_s} = \ln(1 + e_{i_s})$ and by setting $b_{i_0,i_1,\ldots,i_S} = \ln(\mathbb{E}[L_{i_0,i_1,\ldots,i_S}]/\hat{LR})$ the problem can be rewritten as linear programming problem which can be efficiently solved by standard software tools.

$$\begin{aligned} \min \sum_{i_0\in\mathcal{I}_0} u_{i_0} + \sum_{i_1\in\mathcal{I}_1} u_{i_1} + \cdots + \sum_{i_S\in\mathcal{I}_S} u_{i_S} & \\ \text{s.t.} & \\ u_{i_0} + u_{i_1} + \cdots + u_{i_S} &\geq b_{i_0,i_1,\ldots,i_S},\ (i_0, i_1, \ldots, i_S) \in \mathcal{I}, \\ u_{i_1} + \cdots + u_{i_S} &\leq \ln(1 + r^{max}), \\ u_{i_1}, \ldots, u_{i_S} &\geq 0. \end{aligned} \tag{17}$$

		Overd. Poisson			**Gamma**			**Inv. Gaussian**		
Param.	Level	Est.	Std.Err.	Exp	Est.	Std.Err.	Exp	Est.	Std.Err.	Exp
tariff group	1	-3.096	0.042	0.045	10.30	0.015	29 778	10.30	0.017	29 765
tariff group	2	-3.072	0.038	0.046	10.35	0.013	31 357	10.35	0.015	31 380
tariff group	3	-2.999	0.037	0.050	10.46	0.013	34 913	10.46	0.015	34 928
tariff group	4	-2.922	0.037	0.054	10.54	0.013	37 801	10.54	0.015	37 814
tariff group	5	-2.785	0.040	0.062	10.71	0.014	44 666	10.71	0.017	44 679
region	1	0.579	0.033	1.785	0.21	0.014	1.234	0.21	0.016	1.234
region	2	0.460	0.031	1.583	0.11	0.013	1.121	0.11	0.014	1.121
region	3	0.205	0.032	1.228	0.06	0.013	1.059	0.06	0.015	1.058
region	4	0.000	0.000	1.000	0.00	0.000	1.000	0.00	0.000	1.000
age	1	0.431	0.027	1.539	-	-	-	-	-	-
age	2	0.245	0.024	1.277	-	-	-	-	-	-
age	3	0.000	0.000	1.000	-	-	-	-	-	-
gender	1	-0.177	0.018	0.838	-	-	-	-	-	-
gender	2	0.000	0.000	1.000	-	-	-	-	-	-
Scale		0.647	0.000		13.84	0.273		0.002	0.000	

Table 2 Parameter estimates of GLM

5 Stochastic programming problems for rate estimation

In this section, we propose stochastic programming formulations which take into account compound distribution of random losses not only its expected value. We employ chance constraints for satisfying the constraints (9). However, chance constrained problems are very computationally demanding in general, see [3, 4, 5, 6, 11] for various solution approaches and possible reformulations.

If we prescribe a probability level ε for violating the prescribed loss ratio in each tariff cell, we obtain the following chance (probabilistic) constraints

$$P(L_{i_0,i_1,\ldots,i_S} \leq \hat{LR} \cdot Pr_{i_0} \cdot (1+e_{i_1}) \cdot \cdots \cdot (1+e_{i_S})) \geq 1-\varepsilon, \tag{18}$$

which can be rewritten using quantile function of $L_{i_0,i_1,\ldots,i_S}$ as

$$\hat{LR} \cdot Pr_{i_0} \cdot (1+e_{i_1}) \cdot \cdots \cdot (1+e_{i_S}) \geq F^{-1}_{L_{i_0,i_1,\ldots,i_S}}(1-\varepsilon) \tag{19}$$

Setting $b_{i_0,i_1,\ldots,i_S} = \ln[F^{-1}_{L_{i_0,i_1,\ldots,i_S}}(1-\varepsilon)/\hat{LR}]$ formulation (17) can be used. However, it can be very difficult to compute the quantiles for compound distribution, see [16]. Instead of approximating the quantiles, we can employ one-sided Chebyshev inequality based on the mean and variance of the the compound distribution resulting in the constraints

$$P(L_I \geq Pr_I) \leq \frac{1}{1+(Pr_I-\mu_I)^2/\sigma_I^2} \leq \varepsilon, \text{ for } Pr_I \geq \mu_I, \tag{20}$$

which can be rewritten as

$$\frac{1-\varepsilon}{\varepsilon}\sigma_I^2 \leq (Pr_I-\mu_I)^2. \tag{21}$$

This leads to the following reliability constraints

$$\mu_I + \sqrt{\frac{1-\varepsilon}{\varepsilon}}\sigma_I \leq Pr_I. \tag{22}$$

Setting $b_I = \ln[(\mu_I + \sqrt{\frac{1-\varepsilon}{\varepsilon}}\sigma_I)/\hat{LR}]$ we can employ linear programming formulation (17).

		GLM		**MV-model**		**SP-model I**		**SP-model II**	
	Level	G	IG	G	IG	G	IG	G	IG
tariff group	1	1 880	1 879	3 881	3 877	127 164	253 227	7 916	12 112
tariff group	2	2 028	2 029	4 186	4 187	135 565	277 303	8 483	13 209
tariff group	3	2 430	2 431	5 017	5 016	156 748	337 577	9 976	16 003
tariff group	4	2 840	2 841	5 863	5 862	176 535	394 915	11 437	18 715
tariff group	5	3 850	3 851	7 948	7 946	223 966	542 834	14 993	25 627
region	1	2.203	2.201	.241	.240	.464	.470	.293	.358
region	2	.775	.776	.000	.000	.250	.200	.077	.105
region	3	.301	.299	.000	.000	.037	.000	.000	.000
region	4	.000	.000	.000	.000	.000	.000	.000	.000
age	1	.539	.539	.350	.351	.248	.244	.363	.316
age	2	.277	.277	.121	.121	.133	.132	.188	.166
age	3	.000	.000	.000	.000	.000	.000	.000	.000
gender	1	.000	.000	.000	.000	.000	.000	.000	.000
gender	2	.194	.194	.194	.194	.095	.094	.135	.119

Table 3 Optimal rates and segmentation coefficient

6 Numerical example

In this section, we apply proposed approaches to Motor Third Party Liability contracts. We consider 60000 policies which are simulated using characteristics of real MTPL portfolio of one of the leading Czech insurance companies. The following criteria are used in GLM as the independent variables:
1. tariff group: 5 categories (up to 1000, over 1350, over 1850, over 2500, over 2500 ccm engine),
2. age: 3 categories (18-30, 30-65, 65 and more years),
3. region: 4 categories (over 500 000, over 50 000, over 5 000, up to 5 000 inhabitants),
4. gender: 2 categories (men, women).

We employ the approaches proposed in the previous sections to find the basic premium levels for the tariff groups and the surcharge coefficients for other criteria. The goal loss ratio for new business is set to 0.6 and the maximum feasible surcharge to 100 percent. The parameter estimates for overdipersed Poisson, Gamma (G) and Inverse Gaussian (IG) generalized linear models can be found in Table 2. Standard errors and exponentials of the coefficient are also included. All variables are significant based on Wald and likelihood-ratio tests. The parameters of GLM were estimated using SAS GENMOD procedure [17] and the optimization problems were solved using SAS OPTMODEL procedure [18].

The basic premium levels and surcharge coefficients can be found in Table 3. It is not surprising that the coefficients which are estimated directly from GLM do not fulfill the business requirements and the highest possible surcharge is much higher than 100 percent. This drawback is removed by the optimization problems. The decrease of the surcharge coefficient leads to the increase of the basic premium levels. We refer to the problem where the expected loss is covered as MV-model. Inappropriate increase of rates can be observed if we use directly the stochastic programming formulation with the reliability type constraints based on Chebyshev inequality with $\varepsilon = 0.1$ (SP-model I). This increase is reduced in the second stochastic programming problem (SP-model II), where lower "weights" (0.1) are assigned to the variance in formula (22). Stochastic programming models with Inverse Gaussian regression lead to higher estimates of the basic premium levels because the estimated variance is much higher then for Gamma regression. Thus, the first model leads to safer estimates however the variance observed in practice corresponds rather to Gamma model.

7 Conclusion

In this paper, we compared several methods for rating of non-life (MTPL) insurance contracts which take into account riskiness of various segments. The probability distribution of losses was described by generalized linear models. Direct application of the estimated coefficient leads to the surcharge coeffi-

cients which do not fulfill the business requirements. Therefore, optimization models were introduced. Stochastic programming formulation was employed to consider the distribution of the random losses on a policy.

Acknowledgements

This work was supported by the Czech Science Foundation under grant [P402/12/G097].

References

[1] Antonio, K., and Beirlant, J.: Actuarial statistics with generalized linear mixed models. *Insurance: Mathematics and Economics* **40** (2007), 58–76.

[2] Branda, M.: Local stability and differentiability of the Conditional Value at Risk defined on the mixed-integer loss functions. *Kybernetika* **46**(3) (2010), 362–373.

[3] Branda, M.: Stochastic programming problems with generalized integrated chance constraints, Accepted to *Optimization*, 2012. DOI: 10.1080/02331934.2011.587007

[4] Branda, M.: Chance constrained problems: penalty reformulation and performance of sample approximation technique. *Kybernetika* **48**(1) (2012), 105–122.

[5] Branda, M.: Sample approximation technique for mixed-integer stochastic programming problems with several chance constraints. *Operations Research Letters* **40**(3) (2012), 207–211.

[6] Branda, M., and Dupačová, J.: Approximations and contamination bounds for probabilistic programs. *Annals of Operations Research* **193**(1) (2012), 3–19.

[7] Branda, M., and Kopa, M.: DEA-Risk Efficiency and Stochastic Dominance Efficiency of Stock Indices. *Czech Journal of Economics and Finance* (Finance a uver) **62**(2) (2012), 106-124.

[8] Bühlmann, H., and Gisler, A.: *A course in credibility theory and its applications*. Springer Science & Business, 2005.

[9] Denuit, M., Marchal, X., Pitrebois, S., Walhin, J.-F.: *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*. John Wiley & Sons, Chichester, 2007.

[10] Dupačová, J., and Kopa, M.: Robustness in stochastic programs with risk constraints. *Annals of Operations Research* (Online first), 2012. DOI: 10.1007/s10479-010-0824-9.

[11] Houda, M., and Kaňková, V.: Empirical Estimates in Economic and Financial Optimization Problems. *Bulletin of the Czech Econometric Society* **19**(29) (2012), 50–69.

[12] de Jong, P., and Heller, G.Z.: *Generalized Linear Models for Insurance Data*. Cambridge University Press, New York, 2008.

[13] Lachout, P.: Approximative solutions of stochastic optimization problems. *Kybernetika* **46**(3) (2010), 513–523.

[14] McCullagh, P., and Nelder, J.A.: *Generalized Linear Models*. 2nd Ed. Chapman and Hall, London, 1989.

[15] Ohlsson, B.: Combining generalized linear models and credibility models in practice. *Scandinavian Actuarial Journal* **4** (2008), 301–314

[16] Withers, Ch., and Nadarajah, S.: On the compound Poisson-gamma distribution. *Kybernetika* **47**(1) (2011), 15–37.

[17] SAS/STAT 9.3: User's Guide.

[18] SAS/OR 9.3 User's Guide: Mathematical Programming.

AHP analysis of teacher's managerial competencies

Helena Brožová[1]

Abstract. The students of the Czech University of Life Sciences, Prague had to identify and evaluate their expectance of teacher's competencies. The aim of this research is not to evaluate the teacher's scientific ability which can't be subject of students' evaluation. The key characteristics of the managerial competencies of teachers are set according to the Casselmann typology of teacher's roles. Then the students' pairwise comparisons of various teacher's characteristics and competencies were analysed using the Analytic Hierarchy Process.

Keywords: AHP method, teacher's managerial competencies, student's preferences.

JEL Classification: C44
AMS Classification: 90B50

1 Introduction

This article deals with the methodology of identification of the students' preferences of teacher's managerial competencies at the Czech University of Life Sciences, Prague (CULS Prague). The goal is not to evaluate the teacher's scientific ability but the advantage of the teacher's managerial competencies given by students.

Teachers usually think that students are receiving and understanding information in the same way as teachers give it (Skarupská 2007). But do the teachers know what students expect, which pedagogical methods they prefer, what they want not from scientific but from organizational point of view? Therefore the aim of this article is to describe the methodology of how to identify the student's preferences of teacher's managerial competencies.

Preferences could be described as an individual's regard to a set of objects typically in decision-making process (Lichtenstein & Slovic, 2006). Alternatively, preferences mean evaluative judgment as liking or disliking an object (Scherer, 2005). Preferences are generally set as weights. For evaluation of these weights there are many different methods that varied in the proportion of including the subjective and objective judgement. Commonly diffused method is the Analytic Hierarchy or Network Process by T. Saaty (AHP or ANP). The AHP method is a method deriving global weights from partial weights received as result of pairwise comparisons (Saaty, 1980, 1999).

To evaluate teacher's managerial competencies in complexity, we proposed the questionnaires for pairwise comparisons of various teacher's managerial characteristics and competencies. Student's answers are then analysed using the AHP method.

2 Method and Data

2.1 Model Structure

It is very difficult to evaluate managerial competencies of teachers in complexity; we excluded the technical competencies of teacher from observation. For the rest of managerial competencies of teachers were found the key characteristics from the student's point of view. The base for identification of teacher's managerial competencies had been the Casselmann typology of teacher's roles, which was disintegrated to lower levels (Casselmann, 1967). These levels came from managerial competencies (Koontz et al, 1980) and were described according to Philip Morris competencies model (Hroník, 2006). The competencies observed in the study are in **Table 1** (Brožová et al, 2011, Brožová, 2011).

In **Table 1** the teacher's competencies are organised into three groups and it is possible to create the hierarchy of this competency system. And more, it is possible to suppose, that the students preferences differ according to the intensity of the competency characteristic.

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

Competencies groups	Competencies	Characteristics/Anti-characteristics
• L21 Content and form of teaching	• L311 Amount of information	• L411High/Low amount of information
	• L312 Complexity of reading	• L412 High/Low complexity of reading
	• L313 Content of reading	• L413 Oriented on the form of reading/Oriented on the content of reading
	• L314 Form of reading	• L414 Oral/IT based presentation
	• L315 Depth of reading	• L415 Narrow specialization/Broad overview
	• L316 Way of reading	• L416 Innovative/Classical education methods
• L22 Organisation of lecture	• L321 Focus on group or individual	• L421 Individual/Group focus
	• L322 Setting the rules	• L422 Consistent/Changeable decision making
	• L323 Solving problems	• L423 First hand/Diplomatic manner
	• L324 Evaluation process	• L424 Quantitative/Qualitative evaluation methods
	• L325 Evaluation criteria	• L425 Consistent/Changeable criteria
	• L326 Plan of teaching	• L426 Fixed/Framework education plan
	• L327 Flexibility	• L427 Impressible/Uninfluenced
	• L328 Monitoring	• L428 Follow/Do not follow control or monitoring
• L23 Personality of teacher	• L331 Teacher's self-presentation	• L431 Quiet/Vigorous way of speaking
	• L332 Communication skills	• L432 Good/Poor communication skills
	• L333 Focus on student	• L433 Students/Topic orientation
	• L334 Support of student's independence	• L434 Directive/Democratic manner
	• L335 Ability to improvise	• L435 React/Do not react to students
	• L336 Teacher's outlook	• L536 Casual/Informal look
	• L337 Way of speaking	• L437 Professional/Conversional language style

Table 1 Competencies groups and their elements

The whole competency system is really complicated and comprehensive and preference information can have many different forms; therefore its transformation into numerical expression is necessary for mathematical models calculation. So students' weights of these teacher's competencies are estimated as preferences received using Saaty pairwise comparisons methods and subsequently synthesized using the AHP method. The AHP method using quantitative pairwise comparisons is the suitable tool for this analysis, because it enables above described evaluation by sequential comparisons of all possible pairs of items. The AHP is a method deriving global preferences from partial preferences that represent relative measurements of the hierarchical dependences of decision elements (Saaty, 1980, 1999). Fundamental characteristics of both methods are following.

2.2 Analytic Hierarchy Process (AHP)

The AHP (Saaty 1980, 1999) is based on mathematics and psychology. The procedure for using the AHP consists of the following steps:

i. Creation of the problem hierarchy containing the decision goal, the variants for reaching it, and the criteria for evaluating the variants.

   The AHP model for setting of preferences has four levels (**Table 1**): the first one L1 consists of the goal – the preference setting, the second L2x comprise the groups of competences, the third L3xx includes the competencies and the fourth L4xx consists of qualitative characteristics describing the competencies.

ii. Calculation of the priorities among the elements of the hierarchy by making a series of judgements based on pairwise comparison of the elements.

   Pairwise comparison is the process of comparing pairs of items to judge which element of each pair is preferred, or has a greater amount of some quantitative property. One broadly used method is Saaty's pairwise comparison method (Saaty, 1980).

iii. Checking the consistency of the judgements.

iv. Synthesis of these judgements to yield a set of overall priorities for the hierarchy.

v. Selection of the best variant based on the highest overall priority.

2.3 Questionnaire and Respondents

Questionnaire

To receive the necessary data for this analysis, the student's survey was made. The students filled the questionnaire in MS Excel (**Figure 1**) and then the answers were synthesized by the AHP for every questionnaire (Brožová 2011). These data then are worked up using MS Excel tools – functions and also macros.

A	Strong prefered A			Equal preferences					Strong prefered B		B
	9	7	5	3	1	3	5	7	9		
Content and form of teaching					X						Organisation of lecture
Content and form of teaching						X					Personality of teacher
Organisation of lecture						X					Personality of teacher

Figure 1 The part of questionnaire

Respondents

The pilot study was done for really small group of 4 students and this study showed the reasonability of this approach (Brožová 2011). Then the research was done for three groups of students of the last course in Master programs in two faculties of CULS Prague:

- Faculty of Environmental Science
  - 53 regular students of Engineering Ecology, and Landscape Engineering
- Faculty of Economics and Management
  - 48 regular students of Economics and Management, and Administration and Management
  - 127 distance students of Economics and Management, and Administration and Management

Together 228 responses of the student of our university were analysed.

2.4 Processing of Questionnaires

All questionnaires were firstly checked for completeness and were found missing answers. In these cases the equal preference was added.

Saaty's matrices were recalculated automatically using sheets functions and consistency index was calculated using Goal seeking. Then the consistency was checked and inconsistent answers were discarded. The individual weights were calculated using sheets functions at the end of this step. **Figure 2** shows sheet organisation for Saaty's matrix calculation, consistency index checking and weights calculation for competencies and competency groups. Weights of characteristics and anti-characteristics are calculated as shown in **Figure 3**.

9,00	7,00	5,00	3,00	1,00	0,33	0,20	0,14	0,11		Content and form of teaching	Organisation of lecture	Personality of teacher	Geommean		Weights
0	0	0	0	1	0	0	0	0	Content and form of teaching	1,000	1,000	0,333	0,6933613		**0,2**
0	0	0	0	0	1	0	0	0	Organisation of lecture	1,000	1,000	0,333	0,6933613		**0,2**
0	0	0	0	0	1	0	0	0	Personality of teacher	3,000	3,000	1,000	2,0800838		**0,6**
													3,46680637		
										-2,000	1,000	0,333	Lambda	3,00005	**Consistency Index**
										1,000	-2,000	0,333	Determinant	-0,0004	**0,00**
										3,000	3,000	-2,000			

Figure 2 Saaty's matrix calculation and consistency index checking

The processing of data was finalised using special macros made for bringing together of all consistent results. When value of consistency index was very bad, corresponding answers are removed from the final elaboration. Because no student can be preferred more than other one, the average weights of competency groups or characteristics were calculated and analysed at the end.

Excel graphs were used for graphical representation of global student preferences of teachers' managerial competencies. Various orders of elements were used for easier interpretation of global preferences. The weights of the groups of competencies and the weights of competencies are ordered from the highest value. The weights of characteristics are ordered in two ways. The first way is based on differences between weight of characteristics and weight of anti-characteristics, the second one is based on maximal value of corresponding weights (weights of characteristic and anti-characteristic).

Strong prefered A			Equal preferences					Strong prefered B	A / B												
9	7	5	3	1	3	5	7	9				9	7	5	3	1	3	5	7	9	
			X						High/Low amount of information	**0,75**	3	0	0	0	1	0	0	0	0	0	**0,25**
							X		High/Low complexity of reading	**0,125**	0	0	0	0	0	0	0	1	0	7	**0,875**
	X								Orientated/Nonoriented on the form of reading	**0,875**	7	0	1	0	0	0	0	0	0	0	**0,125**
						X			Quiet/Vigorous way of speaking	**0,16667**	0	0	0	0	0	0	1	0	0	5	**0,83333**
X									Good/Poor communication skills	**0,9**	9	1	0	0	0	0	0	0	0	0	**0,1**
		X							Oral/IT besed presentation	**0,83333**	5	0	0	1	0	0	0	0	0	0	**0,16667**
			X						Students/Topic orientation	**0,75**	3	0	0	0	1	0	0	0	0	0	**0,25**
							X		Narrow specialization/Broad overview	**0,125**	0	0	0	0	0	0	0	1	0	7	**0,875**
			X						Individual/Group fokus	**0,75**	3	0	0	0	1	0	0	0	0	0	**0,25**
	X								Innovative/Classical education methods	**0,875**	7	0	1	0	0	0	0	0	0	0	**0,125**
X									First hand/Diplomatic manner	**0,9**	9	1	0	0	0	0	0	0	0	0	**0,1**
	X								Consistent/Changeable criteria	**0,875**	7	0	1	0	0	0	0	0	0	0	**0,125**
			X						Consistent/Changeable decision making	**0,75**	3	0	0	0	1	0	0	0	0	0	**0,25**
							X		Quantitative/Qualitative evaluation methods	**0,125**	0	0	0	0	0	0	0	1	0	7	**0,875**
					X				Fixed/Framwork education plan	**0,25**	0	0	0	0	0	1	0	0	0	3	**0,75**
	X								Impressible/Uninfluenced	**0,875**	7	0	1	0	0	0	0	0	0	0	**0,125**
								X	Directive/Democratic manner	**0,1**	0	0	0	0	0	0	0	0	1	9	**0,9**
			X						Follow/Do not follow control or monitoring	**0,75**	3	0	0	0	1	0	0	0	0	0	**0,25**
	X								React/Do not react to students	**0,875**	7	0	1	0	0	0	0	0	0	0	**0,125**
				X					Casual/Informal look	**1**	0	0	0	0	1	0	0	0	0	0	**1**
		X							Professional/Conversional language styl	**0,83333**	5	0	0	1	0	0	0	0	0	0	**0,16667**

Figure 3 Weights calculation of characteristics and anti-characteristic

3 Results and Discussion

Preferences of the competency groups

Preferences of competency groups show (**Table 2**), that for students Organisation of lectures is not very important. It can be explained by student's ability to accept changes.

Content and form of teaching	Personality of teacher	Organisation of lecture
0,44884	0,33304	0,21811

Table 2 Preferences of groups of competencies

Preferences within the competency groups

The weights of elements in the level **Content and form of teaching** are in **Table 3**. Surprisingly these results show that for students the Way of reading is much more important than the Complexity of reading and Amount of information.

Way of reading	Form of reading	Content of reading	Depth of reading	Complexity of reading	Amount of information
0,31816	0,22186	0,199	0,12601	0,07065	0,06456

Table 3 Preferences of competencies in Content and form of teaching

In the group of competencies **Personality of the teacher** the competencies have small differences among the weights, only the Teachers outlook has very low preference (**Table 4**). It seems that students are really excited and disturbed if the teacher has some inappropriate communication, speech habits and self-presentation or if he is not able to dialogue with students and to improvise.

Communication skills	Way of speaking	Ability to improvise	Focus on student	Support of student's independence	Teacher's self-presentation	Teacher's outlook
0,19276	0,17019	0,17018	0,15675	0,14512	0,12837	0,03663

Table 4 Preferences of competencies in Personality of teacher

Organisation of lectures is not really important for student (**Table 5**) because weights of competencies are not high. In this group of competencies only the Way of solving problems can be mentioned. It seems that students are very interesting and influenced by teacher attitude to problems.

Solving problems	Flexibility	Evaluation criteria	Plan of teaching	Evaluation methods	Monitoring	Setting the rules	Focus on group or individual
0,17154	0,14665	0,13738	0,13697	0,13328	0,11095	0,09999	0,06323

Table 5 Preferences of competencies in Organisation of lecture

The analysis of the synthesised information on the fourth level of hierarchy

Synthesised weights on the fourth level show the preferences of teacher's managerial characteristic (and anti-characteristic) from the quantitative point of view.

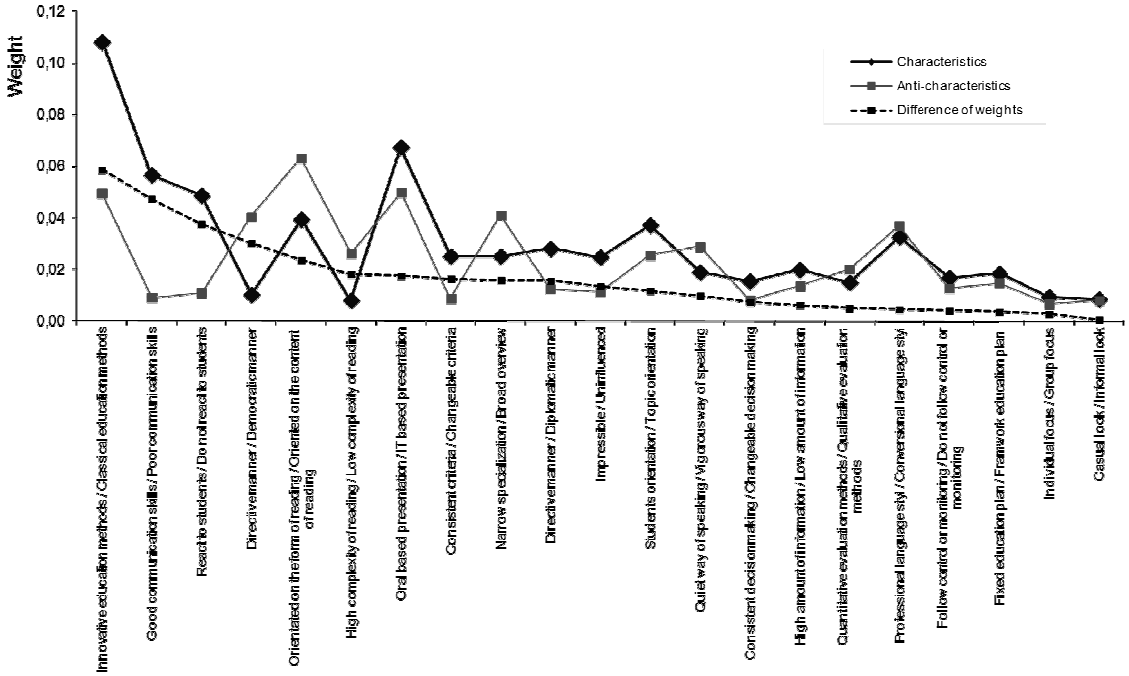


Figure 4 Preferences of managerial characteristic and anti-characteristic (according to the weights differences)

The **Innovative educational methods** (not classical), **Good communication skills**, and immediate **Reaction on students** and to their problems are most preferred by students in contrast with its anti-characteristics. The students want to enjoy their study (**Figure 4**).

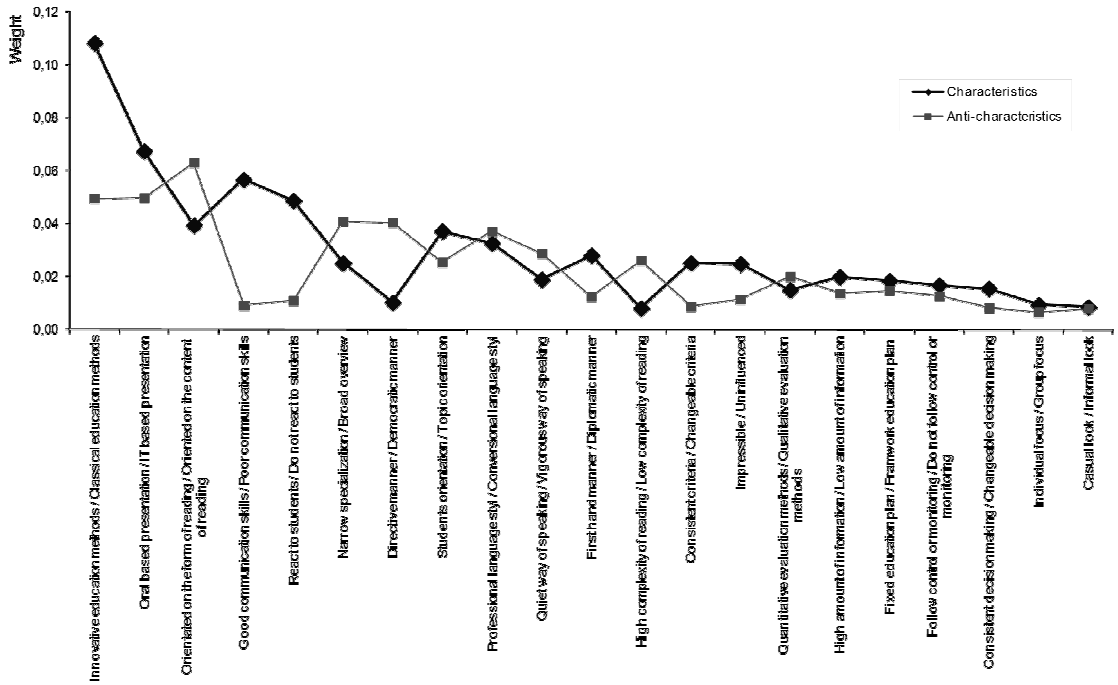


Figure 5 Preferences of managerial characteristic and anti-characteristic (according to the maximal weights)

The **Innovative educational methods** (not classical), **Oral based presentation** (not IT based), and **Orientation to the content of reading** are most preferred by students (**Figure 5**).

4 Conclusion

This article describes the methodology and results of evaluation of students' preferences of teacher's managerial competencies. The questionnaire for evaluation of preferences of competencies and the AHP model was prepared. Questionnaire for students takes not more than 10 minutes of their time, so students are willing to fill them. Totally 228 students were interviewed and results show that for students are much more important:

- Way of reading, Form and Content of reading in the group Content and form of teaching.
- Communication skills, Way of speaking, and Ability to improvise in the group Personality of the teacher
- Way of solving problems in the group Organisation of lecture.
- Innovative educational methods which are more preferred to classical methods.

Acknowledgements

This paper is supported by the project "Interdisciplinary design in the knowledge and skills diffusion process of employees in the Economical Faculty of Czech University of Life Sciences Prague", No. CZ.2.17/1.1.00/33059, supported by the Operational Program Prague – Adaptability and financed with the European Social Fund in the Czech Republic, Czech Republic state budget, and budget of the Capital city of Prague. Prague & EU: We invest to your future.

References

[1] Brožová, H. et al.: Students' Expectance of Teacher's Managerial Competence, In: *Proceedings of the 8th International Conference on Efficiency and Responsibility in Education (ERIE 2011)*, Prague, pp. 36-45, 2011.

[2] Brožová, H.: Weighting of Students' Preferences of Teacher's Competencies, *Journal on Efficiency and Responsibility in Education and Science*, 2011, vol. 4, n. 4, pp. 170-177, ISSN: 1803-1617.

[3] Casselmann, Ch.: *Wesensformen des Lehrers, Versuch einer Typenlehre*, 3. erweiterte Auflage, Stuttgart, Klett, 1967.

[4] Hroník, F.: *Hodnocení pracovníků (Evaluation of employees)*, Praha, Grada, 2006.

[5] Koontz, H., O'Donnell, C., Weihrich, H.: *Management*, 7th ed., McGraw-Hill, New York, 1980.

[6] Lichtenstein, S., Slovic, P.: *The construction of preference*, New York, Cambridge University Press, 2006.

[7] Saaty, T. L.: *The Analytic Hierarchy Process*, New York, Mc Graw Hill Inc, 1980.

[8] Saaty, T. L.: The seven pillars of the analytic hierarchy process, In: *Proceedings of the ISAHP Conference (AHPIC 1999)*, Kobe, pp. 20-33, 1999.

[9] Skarupská, H.: *Styly učení v práci učitele (Teaching styles in teacher's work)*, Ústav pedagogiky a sociálních studií, 2007, [online] Available at: http://www.pedagog.ic.cz/skarupska_dok/soubory/styly_uceni.pdf [Accessed 28 March 2011].

[10] Scherer, K.R.: What are emotions? And how can they be measured? *Social Science Information*, 44, 695-729, 2005.

Agent-based model for comparison of aircraft boarding methods

Richard Cimler[1], Eva Kautzká[2], Kamila Olševičová[3], Martin Gavalec[4]

Abstract. Airline companies need effective boarding methods. Steffen and Hotchkiss compared five boarding methods using real simulation, i.e. experimenting with passengers-volunteers and the realistic model of aircraft. They focused on methods that require arrangement of passengers at the gate. We explored the same boarding methods using agent-based computational model and we have suggested a new method which reduced the boarding time significantly. The influence of the percentage of late or unarranged passengers on the total effectiveness of the boarding process was discussed as well. Our model enables experimenting with various scenarios and parameters' settings such as the number of passengers, the ratio of passengers carrying luggage, the size of the plane etc.

Keywords: agent-based model, aircraft boarding, NetLogo, simulation

JEL Classification: L93, C63
AMS Classification: 90B06, 68M20

1 Introduction

Airports are designed to serve millions of passengers per year. The continuity and accuracy of arrivals and departures is influenced by numerous factors. The efficient boarding procedure is one of them. Agent-based simulation is a useful tool for modelling and exploring dynamic system with interacting individuals. We decided to apply agent-based modelling and simulation to test five boarding methods from [4] together with several new methods we have designed.

The organization of the rest of the paper is as follows. The aircraft boarding methods are analyzed in chapter 2, the NetLogo agent-based model is presented in chapter 3, the experiments are described in chapter 4 and the achieved results are summarized in chapter 5.

2 Aircraft Boarding Methods

The airport boarding is a queuing problem: passengers enter the aircraft one by one, look for their seats, stow their luggage to the box above the seats and sit down. Most airlines use *assigned seating*, i.e. passengers cannot change seat numbers that are printed on their boarding tickets The queue of the passengers in front of the aircraft is not organized (except the preference of the first class, mothers with children, or elderly). Numerous obstructions appear during the boarding process when passengers with the window seats ask passengers from the middle and the aisle seats to get up or multiple passengers put their baggage to the same box etc. These interferences result in delays. The problem of the boarding optimization was analyzed e.g. by Lawson who focused on *unassigned seating* and suggested to use smart swarm concept and ant-based algorithms [2].

Other approaches are based *on arranging passengers* before entering the aircraft. These methods were explored by Steffen and Hotchkiss whose article [4] inspired us. Steffen and Hotchkiss simulated the boarding process with 72 passengers-volunteers and the airplane with 12 rows of six seats and a single central aisle. The objective of optimization was to decrease the number of obstructions, i.e. reduce collision of passengers at the aisle and eliminate the waiting time.

[1] University of Hradec Králové, Faculty of Informatics and Management, Department of Information Technologies, Rokitanského 62, 500 02, Hradec Králové, richard.cimler@uhk.cz.
[2] University of Hradec Králové, Faculty of Informatics and Management, Department of Information Technologies, Rokitanského 62, 500 02, Hradec Králové, kautzka.eva@gmail.com.
3 University of Hradec Králové, Faculty of Informatics and Management, Department of Information Technologies, Rokitanského 62, 500 02, Hradec Králové, kamila.olsevicova@uhk.cz.
[4] University of Hradec Králové, Faculty of Informatics and Management, Department of Information Technologies, Rokitanského 62, 500 02, Hradec Králové, martin.gavalec@uhk.cz.

We focused on 6 methods:

- *Random* – all passengers are boarding together, without specific order. This method corresponds to the traditional boarding process.
- *Wilma* (windows – middle – aisle) – all passengers seated at the windows are boarding in the first group, followed by the middle seats group and the aisle seats group. Inside the group passengers are ordered randomly, therefore there are no seat interferences (situations when aisle or middle seat is occupied before than the window seat), but many aisle interferences.
- *Back-front* – boarding from the back to the front of the aircraft. Passengers seated at the windows are boarding first, followed by the middle and the aisle seats. This method eliminates seat interferences as well as some aisle interferences.
- *Blocks* – boarding in four-row blocks. The back four rows are the first boarding group, followed by the front block and finishing with the center four rows block. The order of passengers in the block is random. Generally, the size of blocks is optional. The optimal number of blocks depends on the length of the aircraft. The number of rows in the block relates the number of seat interferences and aisle obstructions.
- *Steffen* – adjacent passengers in line are sitting two rows apart from each other in corresponding seats (e.g. 12A, 10A, 8A, 6A, etc.). This method attempts to eliminates seat interferences and, as much as possible, aisle interferences while allowing multiple passengers to stow their luggage simultaneously (see Fig. 1).
- *Kautzka 3* – our combination of three principles: *Wilma* and *Back-front* and parallel luggage stowing (see Fig. 1). The method is designed to eliminate both seat and aisle interferences. Our method is also trying to avoid splitting pairs of passengers sitting next to each other – families or colleagues who travel together. More detailed information and our other innovated methods are described in [1].

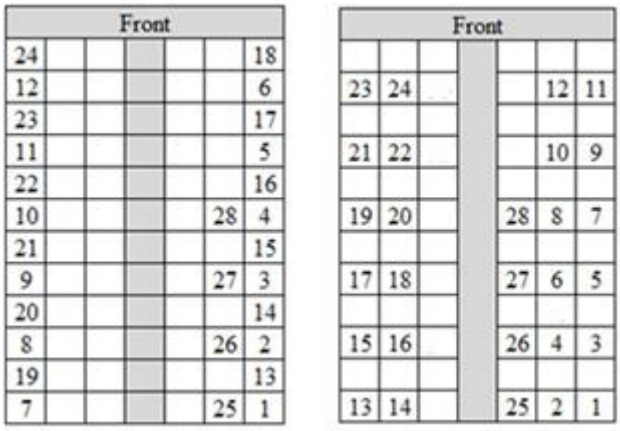


Figure 1: *Steffen* (left) and *Kautzka 3* (right) boarding method

The authors of the *Steffen* method expected that this is the fastest boarding method for airplanes with one entrance and one aisle. Although they did not consider seat interferences, their real simulation results confirmed their expectation. The boarding times were approximately 3.5 minutes for *Steffen* method, over 4 minutes for *Wilma* and *Random*, 6 minutes for *Back-front* and nearly 7 minutes for *Blocks* [4].

Steffen and Hotchkiss could not repeat the real simulation several times and they used only a single setting of parameters (i.e. total number of passengers, number of passengers with/without bags and/or roll-abroad carry-ons). The main problem is that passengers were volunteers and hired actors. Stress, fatigue and other factors does not affect passengers' behavior in the same way as real passengers at airports. Steffen and Hotchkiss' experimental result surely differs from reality in absolute values of boarding times, but the proportion between boarding methods is probably the same also for other settings of parameters.

The boarding methods could be compared in relation to the length of the plane, too. The more rows of seats, the longer boarding process is and the more significant impact of the proper boarding method. In case of the *Steffen* method the total time grows linearly. The *Wilma* method and the *Random* method are less efficient in longer planes because of the unequal distribution of passengers stowing luggage in parallel.

Our objective was to verify the ratios of boarding times from [4]. Moreover we focused on testing the following hypotheses: (a) *Random* boarding is not significantly slower than *Wilma* method, (b) *Blocks* boarding is slower than *Wilma*, *Random* and *Steffen* methods. Here we expected the impact of different usage of the aisle when stowing baggage. *Wilma*, *Random* and *Steffen* methods distribute passengers uniformly along the plane, while in *Blocks* method passengers from the same group obstruct each other.

3 Model in NetLogo

A model is a schematic description of a system, theory, or phenomenon that accounts for its known or inferred properties and may be used for further study of its characteristics. Our agent-based boarding model is built to enable comparing different boarding methods, especially the methods based on specific arrangement of passengers when entering the aircraft. The key element of the agent-based model is an *agent*, autonomous entity with its behavior and properties that operates in the environment. Here the movable *passenger-agents* correspond to passengers, while the plane is represented by *luggage-agents* and the environment. The objective of the agent is to move though the plane to the assigned seat. The sharable description of the model according the standard Overview-Design-Details protocol [3] is presented in [1]. The model was implemented in NetLogo 5.0, a programmable modeling environment well suited for modeling complex systems developing over time [5]. The interface allows modification of parameters of the model and performing experiments (Fig. 2).

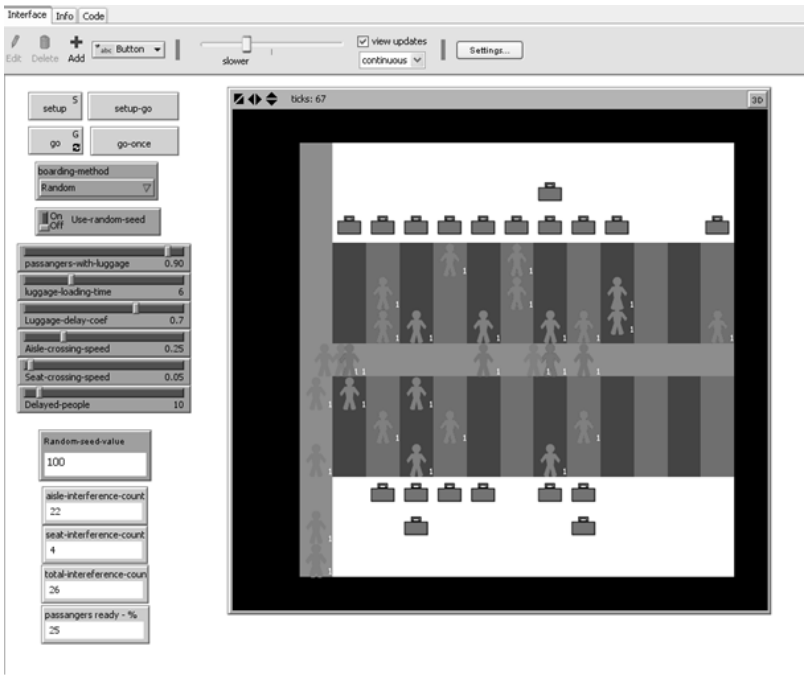


Figure 2: Model interface

Basic assumptions of the model are the same as Steffens' and Hotchkiss' in [4]:

- The plane is fully occupied.
- Each passenger-agent moves independently (the parent-child pairs are not taken into account).
- The proportion of passengers with luggage is optional.
- All passenger-agents with luggage require the same luggage space.
- Two back-to-back passengers can stow the luggage in parallel (the time depends on the value of the luggage-delay parameter setting).

General parameters of the model are:

- *Boarding method* – selection list with 6 possible methods,
- *Use-random-seed* and *Random-seed-value* – switch and slider used for optional repeated runs of simulation the with the same setting of the random number generator,
- *Passengers-with-luggage* – slider for setting the probability of having luggage (for each passenger-agent),
- *Delayed-people* – slider for setting the probability of being late (for each passenger-agent),
- *Luggage-loading-time* – slider for setting initial number of *ticks* necessary for stowing the luggage,
- *Luggage-delay-coef* – slider for setting the increase of waiting time for interferencing passenger-agent,
- *Aisle-crossing-speed* – slider for setting the speed of passenger-agent passing other passenger-agent in the aisle,
- *Seat-crossing-speed* – slider for setting the speed of passenger-agent passing other sitting passenger-agent.

Passenger-agents have got following parameters:

- *Row number*
- *Seat number*
- *Sitting* – true/false value that indicates whether the passenger-agent is already sitting
- *Moving* – true/false value that indicates whether the passenger-agent can move at the moment
- *Waiting* – the number of *ticks* the passenger-agent waits at the same place
- *Luggage* – true/false value that indicates whether the passenger-agent has got a piece of luggage
- *Luggage-time* – the number of *ticks* necessary for stowing the luggage
- *Turn* – the ordering number

The NetLogo environment is the grid of patches. The movement of agent is expressed as modification of [x,y]-coordinate in the grid. In the model the passenger-agent moves zero or one patch per tick of the model clock. The agent tests the patch ahead before making the movement and he cannot access the occupied patch (except passing other stowing or sitting passenger-agent, when these options are allowed).

Four types of interferences of passenger-agents are possible:

- The passenger-agent waits because the other passenger-agent blocks the way.
- The passenger-agent waits because the other passenger-agent puts the luggage to the shared luggage box.
- The passenger-agent has to pass the other passenger-agent stowing luggage.
- The passenger-agent has to pass already sitting passenger-agents.

The run of the simulation stops when all passenger-agents are sitting at their seats. The observed value is the number of *ticks* of the internal clock before the model stops. Implementation details can be found at [1].

There are few simplifications in our model in comparison to [4]. Steffen and Hotchkiss assumed that every passenger need free space (width of one seat on each side) while loading luggage. Their *Steffen* method was designed to be efficient with respect to this assumption. Our model permits directly adjacent *passenger-agents* to stow their luggage with no delay. In case of back-to-back stowing passenger-agents the storing time is extended according to the current parameters' setting.

4 Experiments

Multiple experiments with different combinations of *use-random-seed*, *boarding-method*, *passengers-with-luggage*, *delayed-people* and *luggage-loading-time* parameters were performed. Details can be found at [1]. Here we present two experiments: basic comparison of boarding methods and the measurement of the effect of late passengers.

4.1 Experiment A – comparison of boarding methods

We run the simulation 200 times for every method to eliminate the influence of the random seed generator. The mean time (measured in *ticks* of internal NetLogo clock) is presented in graph (Fig. 3). The *Kautzka 3* method is very fast. This combination *Wilma*, *Back front* and *Steffen* method eliminated all intersections in seats and aisle and it enabled parallel luggage stowing. Moreover it allowed boarding in pairs, without unrealistic splitting families of pairs of passengers.

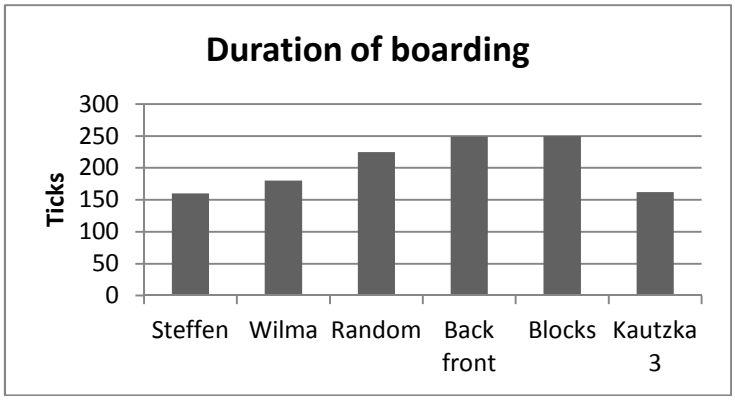


Figure 3: Comparison of all methods

4.2 Experiment B – the impact of late passengers

The disadvantage of *Steffen*, *Back-front* and *Kautzka 3* methods is the expectation of passengers to be punctual and disciplined outside the aircraft. The next experiment was focused on this aspect. We measured the impact of late passengers who could not be arranged into their places in the queue. We tested the influence of 0, 10 and 20 percent late passengers. The results are presented in graph (Fig. 4).

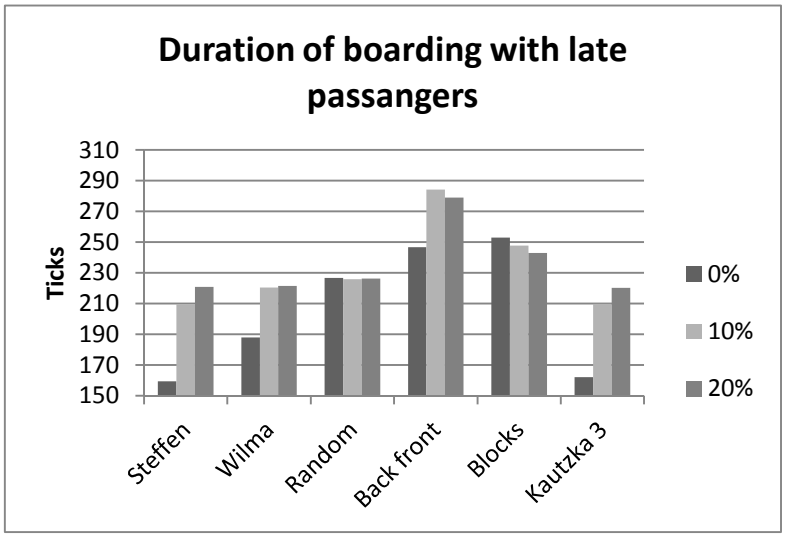


Figure 4: The impact of late passengers

We found that having 10% late passengers, the boarding times prolonged significantly. The *Steffen* and *Kautzka 3* methods are very sensitive to late passengers. Methods *Wilma* and *Back-front* are not so sensitive but there is still increase of average 15%. As we can expect number of late passengers does not change results of the *Random* method. Surprisingly the boarding time for *Blocks* method decreased if number of late passengers increased. The increasing number of late passengers gradually changes the *Blocks* method into the *Random* method which has better results.

Significant increase of boarding time for *Steffen* and *Kautzka 3* methods is caused by increasing number of aisle and seat interferences. Already 10% of late passengers worsen results of *Steffen* and *Kautzka 3* nearly to the *Random* method boarding time.

5 Conclusion

Agent-based modelling is a powerful tool for the analysis of aircraft boarding methods. Our NetLogo implementation of boarding model can be extended easily. It is possible to add new boarding methods as well as to change the size of the plane. We confirmed results of Steffen and Hotchkiss. Moreover we presented an innovated method *Kautzka 3* which improves the Steffen's method by managing groups of passengers who travel together (families, colleagues). Theoretically the usage of sophisticated boarding methods can shorten aircraft preparation for departure. Practically the requirement of precise arrangement of passengers at the gate does not seem to be

realistic because passengers are not used to it. Designing of serviceable methods of arranging passengers could be objective of further research.

Acknowledgements

The research described in the paper was financially supported by grant GACR-402/09/0405.

References

[1] Kautzká, E.: *Agentové simulace hromadné obsluhy*. Diploma thesis, FIM UHK (in Czech), 2012.
[2] Miller, P.: *Smart Swarm Using Animal Behaviour to Change Our World*. Collins, 2011.
[3] Railsback S.F., Grimm, V.: *Agent-based and Individual-based modeling: A Practical Introduction.* Princeton University Press, 2012.
[4] Steffen, J.H., Hotchkiss, J.: Experimental test of airplane boarding methods. *Journal of Air Transport Management*. 18 (1), pp 64-67, 2012.
[5] Wilensky, U.: NetLogo. http://ccl.northwestern.edu/netlogo/ Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL, 1999.

Network structures of the European stock markets

Martin Cupal[1], Oleg Deev[2], Dagmar Linnertová[3]

Abstract. The paper examines changing topological characteristics of correlation-based network of European stock markets on both national and supranational levels. First, the problem of how to correctly build a representative correlation-based procedure and choose a specific filtering procedure for identifying the strongest links is addressed. Then, network structures are investigated on several datasets, for which the data of different time intervals and varying frequency are assembled. On a national level, core stem of stock markets of highly developed countries is found to be stable over time with French market playing the central role. On the supranational level, stocks are clustered based on their economic sector, rather than country's origin. Network modeling of a stock market proves to be highly useful and powerful tool, since network formulation could give much insight and understanding on mutual dependence of stocks' behavior by simply examining graphic representation of the market.

Keywords: stock markets, cross-correlation networks, network topology
JEL Classification: G15
AMS Classification: 91G80

1 Introduction

Economic and financial integration in Europe led to a higher dependability and connectedness of Eurozone stock markets. Developments in financial market of any country – member of the European Monetary Union are perceived by global investors and regulators to highly influence stock markets of other members. But to what degree European stock markets are interconnected and are there any exceptions to the situation? This problem might be addressed by analyzing the network topology of stock markets.

Studies of network properties recently gained a lot of attention from researchers with applications of graph theory widely utilized in biology, sociology, operations research and many other fields of science (for the survey on applications of network theory see [6]). Basic financial analysis of stock markets could also thrive from this approach. For instance, correlation analysis of equity returns in financial markets, usually reported in every study of financial markets in a form of the table of pair cross-correlation coefficients, does not give us a full picture of connectivity between stocks, but, if represented in the form of graph, could give us an interactive and deep understanding of the data for further consideration. The question is how to choose statistically significant correlations and build a network, taking into consideration the full range of information from the dataset.

The aim of our study is to examine changing topological characteristics of correlation-based network of European stock markets on both national and supranational levels. For that purpose the dataset is assembled from country indices and market prices of highly capitalized stocks in different time intervals of varying frequency. The study has a certain practical importance, since it might be utilized for the asset portfolio optimization and the analysis of financial market dynamics.

2 Methodology and Data

In the majority of studies the analysis of network topologies of financial market is described merely as an instrument of ongoing financial market research with no final results reported. To build a network of chosen equity markets or stocks, first, we calculate pair-wise correlations to quantify the degree of synchronization between markets or stocks

[1] Masaryk University, Faculty of Economics and Administration, Department of Finance, Lipová 41a, 60200, Brno, Czech Republic, e-mail: matrass@mail.muni.cz
[2] Masaryk University, Faculty of Economics and Administration, Department of Finance, Lipová 41a, 60200, Brno, Czech Republic, e-mail: oleg@mail.muni.cz
[3] Masaryk University, Faculty of Economics and Administration, Department of Finance, Lipová 41a, 60200, Brno, Czech Republic, e-mail: Dagmar.Linnertova@seznam.cz

and, second, we employ filtering techniques to determine the most important links from the correlation matrix as well as layout algorithms to choose the best way to illustrate the results.

The correlation coefficient for each pair of markets or stocks is defined by:

$$\rho_{ij} = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sqrt{(\langle r_i^2 \rangle - \langle r_i \rangle^2)(\langle r_j^2 \rangle - \langle r_j \rangle^2)}}$$

where *i* and *j* are stock labels, and *r* are market or stock returns (calculated as logarithm price differences).

Next step is to define a metric that clarifies the distance between markets or stocks synchronously evolving in time: $d_{i,j} = \sqrt{2(1 - \rho_{ij})}$ (formula's derivation might be find in [8]). The following three properties (or axioms) must hold:

1) $d_{i,j}$=0 if and only if *i=j;*
2) $d_{i,j}$=$d_{j,i}$;
3) $d_{i,j} \leq d_{i,k} + d_{k,j}$.

A unique way of connection between markets or stocks is specified from the obtained distance matrix by employing the graph theory's concept of minimum spanning tree (MST). In a connected graph *G = (V, E),* each edge *e* is given a weight *w(e)* represented by the calculated metric distance $d_{i,j}$, and weight of a whole graph, which is needed to be minimized, is a sum of weights of edges. Hence, MST is a tree having *n-1* edges that minimize the sum of the edge distances. The problem is how to compute a minimal weighted tree, whose edges cover the entire set of vertices *V*. MST problem is one of the most studied problems in graph theory, for which several solutions or algorithms are known, namely the algorithms of Prim [8], Kruskal [5] and Borůvka [2]. Different filtering procedures could provide different aspects of the time series information. According to several studies (such as [1], [3]), the filtering procedure based on Kruskal's algorithm is a straightforward choice.

The MST associated with the subdominant ultrametric distance matrix *D* can be obtained as follows (here described in spirit of Kruskal [5]). Let assume that the given connected graph *G = (V, E)* is complete, which means that every pair of vertices is connected by an edge. If any edge of *G* is "missing", an edge of greater length may be inserted, and this does not alter the graph in any way relevant to our purpose. Also, it is possible and intuitively appealing to think of missing edges as edges of infinite length. Among the edges of *G* not yet chosen, we pick the shortest edge, which does not form any loops with those edges already chosen. This procedure is performed as many times as possible. Clearly the set of edges eventually chosen must form a spanning tree of *G*, and in fact it forms a shortest spanning tree.

For programming purposes Kruskal's algorithm should be presented as the following procedure:
Step 1. Create an edgeless graph *T = (V, 0)* which vertices correspond with those of *G*.
Step 2. Choose an edge *e* of *G* such that (1) adding *e* to *T* would not make a cycle in *T* and (ii) *e* has the minimum weight *w(e)* of all the edges remaining in *G* that fulfill the previous condition.
Step 3. Add the chosen edge *e* to graph *T*.
Step 4. If *T* spans *G*, procedure is terminated; otherwise, the procedure is repeated from Step 2.

Obtained scale-free graph *T = (V, E')* in a form of a hierarchical tree represents the network of most important correlation-based connections of equity markets or stocks. Vertices or nodes symbolize different time series (or in our case index or stock returns) and are connected by edges or arcs with a weight (thickness of the edge) related to the correlation coefficient between two indices' or stocks' returns.

The majority of empirical studies exploit US market data to investigate network topologies of financial markets. We consider financial market of the Eurozone countries to be a perfect experiment field for our network study, where all usual limitations in the studies of stock markets are not presented. Trading hours of studied stock exchanges are synchronized with the same opening and closing hours (with few exceptions for the smallest stock exchanges). Transactions are made in one currency - euro, so this not imposes additional restrictions on the model specification due to exchange rate fluctuations.

Our empirical analysis is based on four datasets of different time horizons for 17 indices, representing all members of the European Monetary Union (major stock market characteristics are summarized in Table 1):

— one-year daily data of stock indices' prices from April 1st, 2011 till March 30th, 2012;
— intraday 30-minute data of stock indices' prices from March 1st, 2012 till March 21th, 2012;

— intraday 5-minute data of stock indices' prices for March 16th, 2012;
— intraday 30-minute data of prices for 300 highly-capitalized stocks from March 1st, 2012 till March 21th, 2012.

Tick symbol	Country	Number of listed stocks	Market capitalization, mln. US$	Market capitalization as % of GDP	Market turnover, mln. US$	Market turnover as % of GDP	Market liquidity, %
ATX	Austria	73	82 373,8	17,9	38 725,0	12,7	51,6
BEL20	Belgium	158	229 895,9	57,4	107 236,0	23,7	42,9
CYSMMAPA	Cyprus	117	2 853,2	29,5	484,5	2,7	10
TALSE	Estonia	15	1 611,2	11,8	243,3	1,7	12,6
HEX25	Finland	121	143 080,7	49,6	174 349,5	42,8	133,5
CAC	France	893	1 568 729,8	75,3	1 474 235,4	57,3	84,3
DAX	Germany	670	1 184 458,6	43,6	1 758 185,2	42,8	134,5
FTASE	Greece	275	33 648,2	24,1	24 712,0	14,3	46,5
ISEQ20P	Ireland	48	35 362,6	16,3	15 647,3	8,1	42,3
FTSEMIB	Italy	287	431 470,8	15,4	887 454,0	26,2	236,8
LUXXX	Luxemburg	31	67 625,0	189,6	122,7	0,3	0,14
MALTEX	Malta	20	3 424,2	29,1	48,8	0,3	1,7
AEX	Netherlands	108	594.731,6	84,8	554 302,9	76	88,3
PSI20	Portugal	46	61 687,7	35,8	36 143,9	13,7	50,3
SKSM	Slovakia	81	4 736,3	4,8	452,5	0,2	10,2
SBITOP	Slovenia	66	6 325,6	20,1	512,1	0,6	6,5
IBEX	Spain	3241	1 030 951,4	83,3	1 419 228,6	96,7	128,9

Sources: Bloomberg; The World Bank (World Development Indicators)

Table 1 Major stock market characteristics of EMU countries at the end of 2010

Datasets are gained from Bloomberg, where, unfortunately, high-frequency observations for the smallest EMU stock markets (Malta, Luxemburg, Slovakia) are not reported; however, this could not influence the overall analysis. We believe that the analysis of the chosen datasets allows us to capture trading patterns in the most recent market situation from long-term, middle-term and short-term perspective, with the monetary union facing its first stability problem.

3 Results and Discussion

Network is a time-dependent arrangement, but it maintains on a considered time scale a basic structure that exhibits a meaningful economic taxonomy, which is of a main interest to our study. Figures 1 and 2 illustrates the minimum spanning trees of EMU stock markets, obtained by the filtering procedure of pair-wise correlation coefficients of index returns time series computed at 1-year time horizon (with the interval of one day), 3-week time horizon (with the interval of 30 minutes) and 1-day time horizon (with the interval of 5 minutes). Each circle or vertex represents a stock market labeled by its tick symbol used in Bloomberg. Use of different time horizons allows us to investigate modifications of the network's hierarchical organization.

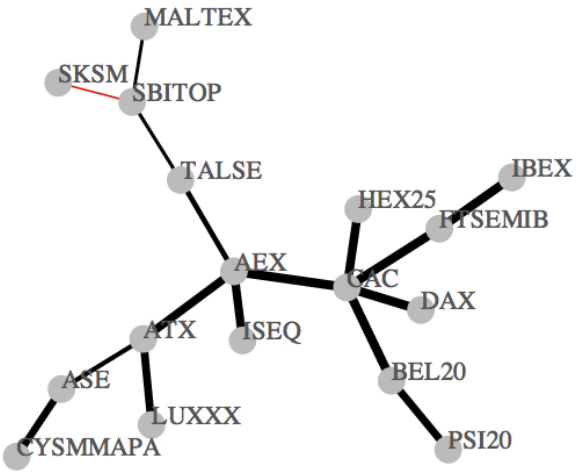


Figure 1 Minimum spanning tree of EMU stock markets from the long-term perspective

Eurozone stock market is perceived as a united market with no stable segmentation. Network's line thickness highlights the significance of four biggest European financial markets: French, German, Italian and Spanish (Table 1), to which all other markets are connected. French stock market plays an unexpectedly central role in the dynamics of the Eurozone financial market, when usually the main attention of investors and regulators is paid to the German market as the representative of the European biggest economy. It does not mean that French market acts as a situation-making player, but because it is the main transmitter of shocks to other markets, it signifies the overall financial situation in the Eurozone.

Markets with small capitalization and lowest market liquidity (Greece, Cyprus, Estonia, Malta, Luxemburg, Slovenia, Slovakia) demonstrate lowest degree of connectivity to other markets. It possibly highlights the illiquidity of European smallest financial markets, which might be concluded as yet another indicator of their inefficiency. Consolidation of such stock exchanges could raise the weight of small economies' financial system in the European context and became an additional impulse for their development (for example, the emergence of Central and Eastern European Stock Exchange Group with the leadership of Vienna Stock Exchange).

Addressing the differences in market topology from different time interval perspectives, we see the stability of its core stem "Frankfurt – Paris – Milan – Madrid" markets, however, appearing in different order with the French market still being a central "hub". Dutch and Belgian stock markets also play a crucial role in the studied group and should not be overlooked in the process of portfolio optimization or policy making.

Evidently, the intensity of market activities on the stock exchange determines the degree of connectivity (statistical significance of the correlation coefficients) of this market to others, when less liquid markets show lesser connectivity due to lesser amount of stocks traded on those markets. Otherwise, it indicates that the dynamics of small stock markets could not match the dynamics of "the biggest four". Thus, to analyze the dependability of the European less-capitalized markets, longer time intervals should be considered.

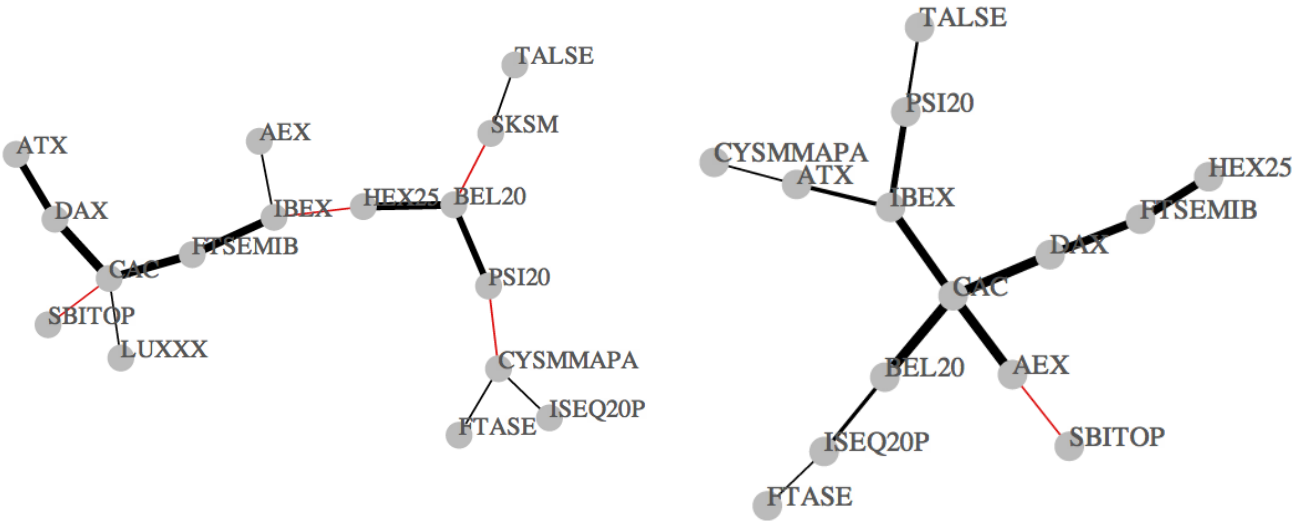


Figure 2 MSTs of EMU stock markets from the middle-term (on the left) and short-term (on the right) perspectives

Analysis of interconnectedness between European stocks sheds the light on whether investors perceive the financial market of the European Monetary Union as a whole or still by its country counterparts. Figure 3 visualizes dependences in a network of about 300 most tradable and highly capitalized stocks in the 17 stock markets, representing all members of the EMU. In most cases, groups of stocks are homogeneous with respect to their economic sector, rather than country's origin. Therefore, the equity market of the European Monetary Union is seen as a truly integrated supranational market. Network of the portfolio of stocks does not coincide with the network of the European stock exchanges.

According to the position of stocks in the network and number of links, the main stem of the tree comprises mainly financial and construction companies (such as Deutsche Bank, Allianz, BNP Paribas, Vinci, Saint-Gobian). Minimum spanning tree clearly exhibits clustering of assets' correlation, where same-class assets are assembled. The biggest clusters, playing the central role in the market dynamics, represent European most developed economic sectors, such as banking, insurance, construction, high-technology, chemical and automobile industries.

Financial network also emphases the anomalies in the time series. Deviations from the observable structure gave valuable information that is displayed in vertices' distancing or their complete detaching. Finnish stocks form the biggest group of distanced vertices. Subtree of Finnish stocks signifies lower degree of connection and integration to other EMU markets, than it was captured by the network of stock indices. This is a clear opportunity for portfolio diversification, supported by the proper market liquidity ratio and growing market capitalization of the Finnish market. As for detached vertices, stocks of illiquid markets previously established, such as Slovakia, Slovenia and Luxemburg, are found on the edge of the tree.

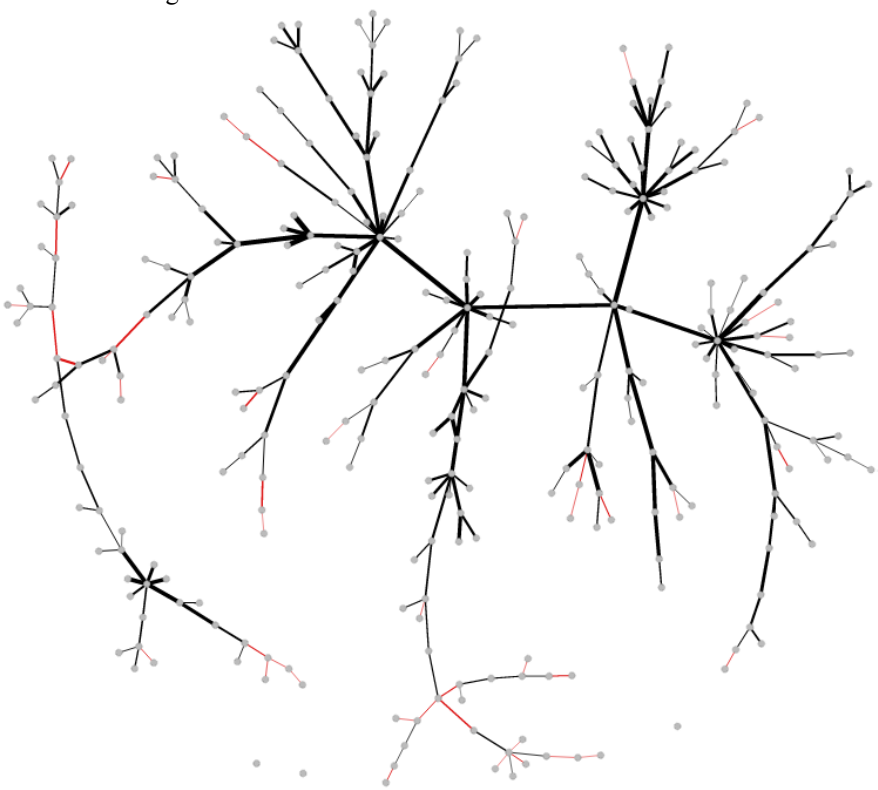


Figure 3 Minimum spanning tree of 300 highly-capitalized European stocks

Investigation of stock networks in other time horizons (not reported here, but available on request) leads to similar results with comparable market segmentations. However, for the purposes of portfolio optimization and policy making network analysis should be conducted on a regular basis.

4 Conclusions

Analysis of stock market topology is a powerful tool to filter meaningful information from correlation coefficient matrix and capture market dynamics, if implemented over time. Network build as a minimum spanning tree allows exploration and monitoring of large-scale dependence structures and dynamics of financial markets in a more interactive way. But we should be also aware of the shortcomings of the approach. The main limitation comes with the sampling time intervals used for the building of the network, which affect the topology of a correlation based network (the problem is deliberately discussed in [1]). On the other hand, this limitation could also be seen as a method of illustrating the complex process of the price formation occurring in financial markets.

For the analysis of the European stock market, the topological properties of the network of stocks should be considered, since it provides deeper understanding and closure to the market structure and connectivity between counties' markets, while also revealing certain market anomalies.

In this paper, the most common method of assets' dependability (cross correlation) was chosen to illustrate the usage of network theory for the analysis of financial markets. However, we believe that more robustness results could be achieved if network structures would be drawn on connections obtained by volatility-based cross correlations or results of the cointegration analysis. Moreover, characteristics of the network topologies could be utilized to validate or falsify widespread managerial models.

References

[1] Bonnano, G., Caldarelli, G., Lillo, F., Micciche, S., Vandewalle, N., Mantgena, R.N. Networks of equities in financial markets. *The European Physical Journal B* **38** (2004), 363-371.

[2] Borůvka, O.: O jistém problému minimálním. *Acta Societatis Scientiarum Naturaliste Moravicae* **3** (1926), 37–58.

[3] Broutin, N., Devroye, L., and McLeish, E.: Note on the Structure of Kruskal's Algorithm. *Journal: Algorithmica* **56-2** (2010), 141-159.

[4] Chartrand, G. *Introductory Graph Theory*. Dover Publications, 1984.

[5] Kruskal, J. B.: On the shortest spanning subtree of a graph and the travelling salesman problem. *Proceedings of the American Mathematical Society* **2** (1956), 48-50.

[6] Lerner, U., Wagner, D., Zweig K.A. *Algorithms of Large and Complex Networks. Design, Analysis and Simulation*. Springer, 2009.

[7] Mantgena, R.N., and Stanley, H.E. *An introduction to econophysics: Correlations and complexity in finance*. Cambridge University Press, Cambridge, 2000.

[8] Prim, R.C.: Shortest connection networks and some generalizations. *Bell System Technology Journal* **36** (1957), 1389–1401.

Comparison of recursive parameter estimation and non-linear filtration

Jan Čapek[1]

Abstract. The contribution compares and contrasts the results of time-varying parameters estimation pursued by two methodologically different approaches, all in the context of a DSGE model. The original results [3] were obtained by the so-called recursive, rolling and "first observation" analysis. The novel methodology uses Unscented Particle Filter to filter trajectories of time-varying parameters.

The comparisons are presented on the case of three parameters of a DSGE model, which demonstrate a match in the evolution of the parameter estimates. Recursive impulse response functions are compared on two cases, which also match.

Although not all of the results obtained from two different methodologies completely match, the most important parameter changes are similarly captured by both approaches.

Keywords: Unscented Particle Filter, recursive analysis, time-varying parameter, DSGE model, impulse response function

JEL classification: C32, C52, E32, E43, E52, F41, F43
AMS classification: 91B51, 91B64

1 Introduction

The contribution compares and contrasts the results of "time-varying parameters estimation" pursued by two methodologically different approaches, all in the context of a DSGE model. The text virtually follows last year's contribution Čapek [3] and verifies its results by an analysis that follows different methodology.

The motivation of the original research was to identify possible structural changes which could demonstrate itself as a changes in (structural) parameters. In order to conduct sensitivity analysis to the method used in the original research [3], this contribution offers a brief verification by fundamentally different state-of-the-art method.

The model – which is not introduced due to the lack of space – is a small-scale Small Open Economy (SOE) New Keynesian Dynamic Stochastic General Equilibrium (DSGE) model with 7 observable variables. For the description of the model, see Lubik and Schorfheide [7] for very similar model.

2 Different methodologies of time-varying parameters estimation

This section offers a brief introduction to two methods that were used to estimate time-varying parameters in the small scale Dynamic Stochastic General Equilibrium (DSGE) model at hand.

2.1 Recursive, rolling and "first observation" analysis

Original results that are to be verified are in Figures 1, 2 and 4. These figures depict results of the so-called "first-observation", recursive and rolling analysis. All of these methods proceed from Bayesian

[1]Department of Economics, Faculty of Economics and Administration, Masaryk University, Lipová 41a, 602 00 Brno, e-mail: capek@econ.muni.cz

estimation of time-invariant parameters. To detect tendencies of fixed parameters to change, a number of estimations are conducted – each on a different sample of the same data set.

For example, adding one extra observation to the data set and re-estimating the whole model could show us that some parameters are different than the original estimates, whereas others are not. Parameter estimates that are still the same with no influence of changing time sample can be considered as time-invariant by nature. On the other hand, if the estimate of a parameter changes as the data sample changes, such parameter is time-variant.

In order to conduct the analyses in a systematic way, there are three types of estimations: the most frequent in the literature[1] is the so-called recursive analysis. This analysis adds one-by-one more data point to the original data sample. After adding each data point, the model is re-estimated and the results stored. This type of analysis therefore detects the changes of parameters connected to the new data.

Second – still frequent in the literature – analysis is the so-called rolling analysis. This analysis deals with the problem of growing size of the data sample by discarding the oldest data point each time a new data point is added to the sample. Rolling analysis is therefore conducted on a sample of fixed length.

The last analysis – that is quite infrequent in the literature – is an analysis that detects changes of the parameters when the oldest data points are discarded. The working title of such analysis is "first observation" analysis since first observations are discarded one-by-one from the data sample.

All in all, the advantages of the described approach are that it detects gradual as well as sudden changes and that it detects changes in any part of the data sample (the beginning and the end included). Also, all of the parameter estimates are consistent with the DSGE system since all of the estimates are the results of a Bayesian estimation. However, there are also drawbacks connected to this approach. The most prominent drawback is that – by nature – each estimation treats parameters as fixed. Therefore, the optimizing agents in the economy are not aware of the time-variant nature of the parameters and in each estimation.

2.2 Nonlinear filtration with Unscented Particle Filter

The new state-of-the-art method that is used to verify the results of a former research in Čapek [3] has an advantage that the parameters are filtered as truly time-variant.

The method also proceeds from Bayesian estimation but then it filters selected parameters as time-variant with Unscented Particle Filter (UPF). See Fernández-Villaverde and Rubio-Ramírez [5] for theoretical background of full-fledged non-linear estimation of time-variable parameters, Tonner et al. [8] for application on the Czech data and Liščinský [6] for the implementation used in this contribution. The time-variant parameters are modelled by the TVP process

$$par_t = \rho_{par} par_{t-1} + (1 - \rho_{par}) par^{SS} + \epsilon_t^{par}, \qquad par_0 = par^{SS} \tag{1}$$

where par_t denotes the evolution of the parameter in time t, $0 > \rho_{par} > 1$ is the persistence parameter of the TVP process, par^{SS} is the steady-state value of the TVP process and ϵ_t^{par} is the shock to the TVP process.

The persistence of the TVP process is controlled by the parameter ρ_{par}. Considering extreme values, if $\rho_{par} = 0$, then the TVP process collapses to $par_t = par^{SS} + \epsilon_t^{par}$ and the process is therefore a noise around the steady-state value with no persistence. For the other extreme, consider $\rho_{par} = 1$ in which case the process is $par_t = par_{t-1} + \epsilon_t^{par}$, which is random walk with no consideration for the steady-state value. The closer is ρ_{par} to 0, the lower is the persistence of the process and the higher is the relevance of the steady-state value.

3 Comparisons

In order to conserve space, this section presents the verification of only those parameter estimates that exhibit the most significant changes as reported in Čapek [3] and also compares the recursive impulse response results that demonstrate interesting changes in model behavior.

[1] See e.g. Canova [1] for the use of recursive and/or rolling analysis and Clarida, Galí and Gertler [2] for what is similar to the so-called "first-observation" analysis.

3.1 Parameter estimates

First example of the comparison concerns parameter h, which is habit persistence in consumption. Figure 1 displays the results of previous research by recursive and rolling analysis. All of the estimates show a decline of the parameters in the beginning of 2009. Therefore, the onset of economic crisis is associated with lower consumption smoothing by households.

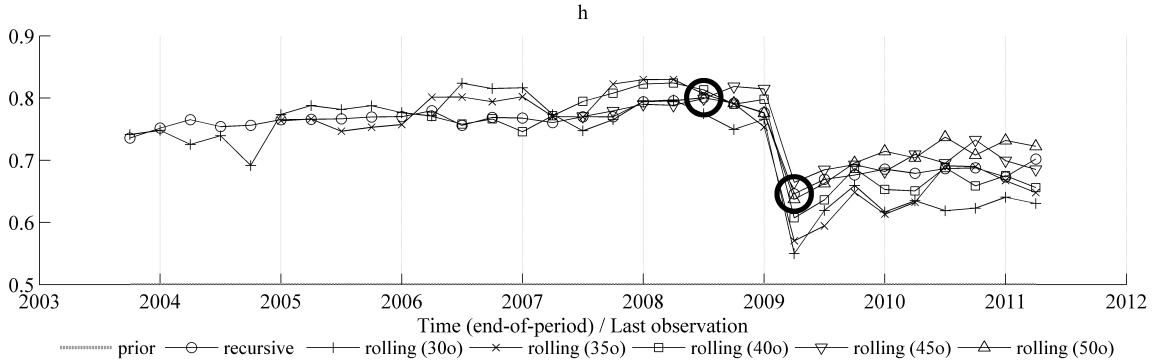


Figure 1: Results of recursive and rolling analysis focused on the end of the series. Case of habit persistence in consumption h.

As for the estimates of h with UPF, is is depicted in the left-hand panel of Figure 3. The estimation with UPF shows a decline in h earlier (than in 2009) and shows also a more significant drop in the parameter. The UPF estimate drops to some 0.45, whereas recursive and rolling estimates (see Figure 1) drop to just 0.65. The fact that households tend to smooth less their consumption during recession was strongly indicated by original analysis Čapek [3] and is concurred by filtration with Unscented Particle Filter.

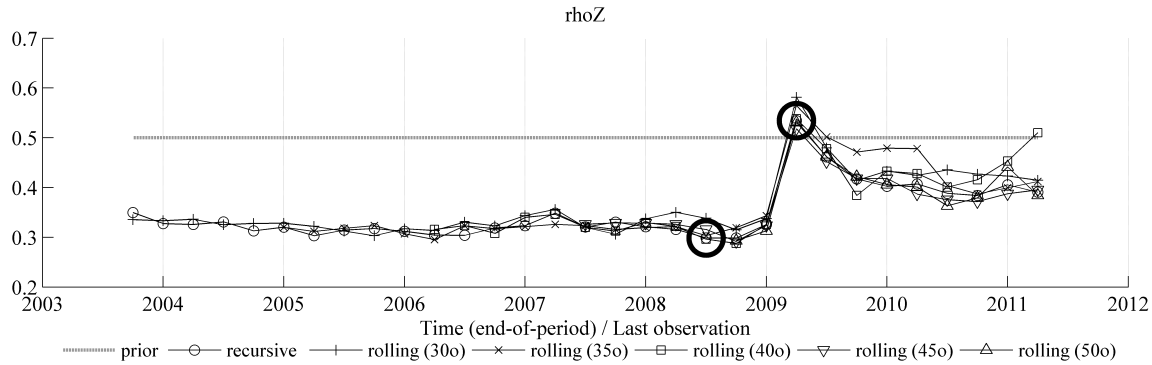


Figure 2: Results of recursive and rolling analysis focused on the end of the series. Case of the persistence in the growth rate of the world-wide technology shock ρ_Z.

The second example concerns a parameter that also changes very prominently according to recursive and rolling analyses. The parameter represents the persistence ρ_Z in the growth rate of the world-wide technology shock z_t such that $z_t = \rho_z z_{t-1} + \epsilon_{z,t}$. The role of the shock is that it hits both model economies in the same manner, i.e. it is symmetric. Greater speed of technology progress therefore helps both economies and, on the other hand, slower speed of technology progress slows both economies down.

As Figure 2 shows, the recursive and rolling estimates exhibit a major rise in the parameter in the beginning of 2009. The parameter jumped from some 0.3 to some 0.5 and then it gradually declines. The estimation of the same parameter with UPF is depicted in the middle panel of Figure 3. Contrary to the previous case of h, the change is far less apparent in the UPF estimation than in the recursive and rolling estimates. The rise in the UPF estimation is "just" from 0.5 to 0.56. However, the fact that the persistence of the growth rate of world-wide technology shock rises in the beginning of the economic crisis is again apparent in both approaches.

Economic interpretation of this parameter change is not trivial.[2] In short, in the time-period in question, which is the beginning of 2009, there were large negative world-wide technology shocks and the increased persistence of its growth rate helped the model to explain the magnitude of the crisis.

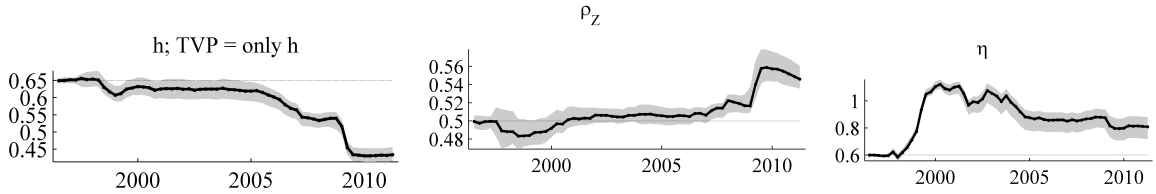


Figure 3: The estimate with UPF of habit persistence in consumption h (left-hand panel), the estimate of the persistence in the growth rate of the world-wide technology shock ρ_Z (middle panel) and the estimate of the elasticity of substitution between home and imported consumption goods η (right-hand panel). Grey area = 95% bands.

The third example addresses a shift in a parameter that was identified in the beginning of the series and the selected parameter is the intratemporal elasticity of substitution between home and imported consumption goods η. Figure 4 shows the so-called "first observation" and rolling estimates. Please note that[3] this figure has the date of first observation on x-axis. Since the analysis is focused on identifying the consequences of discarding data points from the beginning of the series, such depiction is natural. The estimates in Figure 4 show that discarding observations for years 1998–2001 from the data sets unambiguously raises the elasticity of substitution η. In these years, also the estimation with UPF in the right-hand panel of Figure 3 exhibits a rise.[4] In estimates by both methods, the change in the point estimates is quite big: Original analysis (in Figure 4) indicates a rise from 0.2 to almost 1, whereas the filtration with UPF (see Figure 3) shows a rise from steady-state value 0.6 to 1.

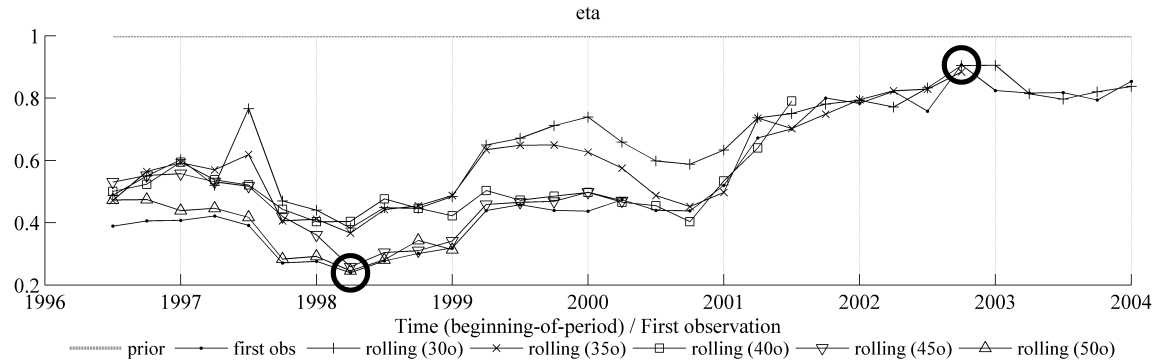


Figure 4: Results of "first observation" and rolling analysis focused on the beginning of the series. Case of intratemporal elasticity of substitution between home and imported consumption goods η.

Although these examples showed that there is broad similarity of the estimates, it is certainly not the case of all parameters. So far, the observation is that for the parameters with major (and statistically significant) changes, both methods concur. In cases of minor changes, the methods report in some cases different results.

3.2 Recursive impulse response analysis

Changing parameter values do not automatically mean changing behavior of economic system since the movements of parameter values may offset each other. Towards this end, it may be interesting to conduct impulse response analysis in a recursive manner to find out if the behavior of the system changes or not.

[2] See Čapek [4] for more in-depth explanation.

[3] in contrast to more usual depiction in Figures 1 and 2.

[4] Please note that Figure 4 and 3 both depict a different time frame but the time span of interest (1998–2001) is in both figures.

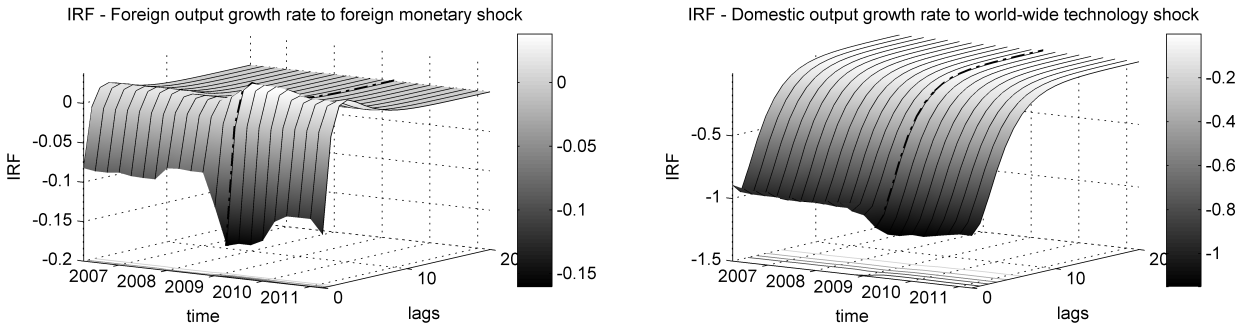


Figure 5: Recursive impulse response functions; parameter values filtered by UPF. Left panel: The shock is to foreign monetary rule (by 0.12). The variable where the shock is observed is foreign output growth. Right panel: The shock to the growth rate of the world-wide technology (by -0.81). The variable where the shock is observed is domestic output growth. Highlighted = 1Q2009.

This section presents two selected recursive impulse response functions which are compared. Figure 5 presents results from recursive impulse response analysis that uses parameter values filtered by UPF. Figure 6 displays results for the same shocks on the same variables, but the parameters of the system were estimated by Bayesian techniques during recursive analysis.

First example – placed in the left-hand panels in both figures – is a reaction of foreign output growth to a shock to foreign monetary rule by 0.12. Both figures naturally exhibit a negative reaction of output growth to a monetary shock. Moreover, in both figures there is apparent drop of the impulse response function in first quarter of 2009. Although the persistence of the change is different in the two cases, the evolution of the impulse response functions is very similar considering that the results are acquired form completely different methodologies.

Another example depicts a reaction of domestic output growth as a result of a shock to the growth rate of world-wide technology. Since the shock is negative (by -0.81), the reaction of output growth is naturally negative as well. The dynamics of change is different in this case. Figure 6 displays a drop in the IRF in the first quarter of 2009 and exhibits greater reaction (than pre-1Q2009) even in the subsequent quarters. On the other hand, Figure 5 displays only slightly more severe reaction of the output growth. Although the severity of the changes is different, again, the fact that the crisis period displays greater reaction holds for both approaches.

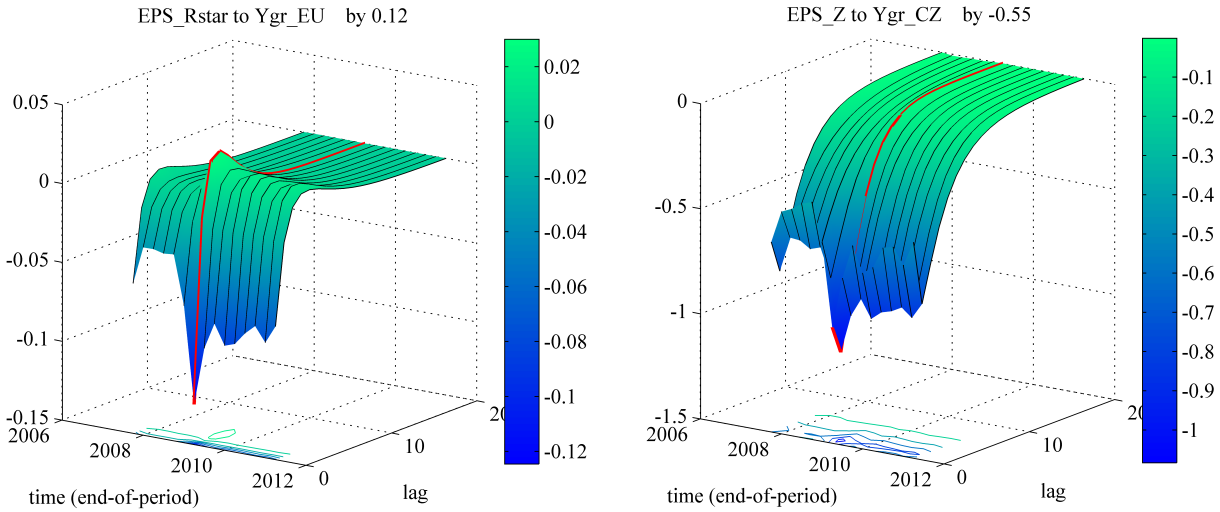


Figure 6: Recursive impulse response functions; parameter values estimated by Bayesian techniques. Left panel: The shock is to foreign monetary rule (by 0.12). The variable where the shock is observed is foreign output growth. Right panel: The shock to the growth rate of the world-wide technology (by -0.81). The variable where the shock is observed is domestic output growth. Highlighted = 1Q2009.

Needless to say, even in recursive impulse response analysis, there are some impulse response functions that behave differently in the two approaches, but the number of similar impulse response functions is fairly high.

4 Conclusion

Due to the limitation to the extent of the contribution, only most prominent results of the comparison were introduced. Section 2 briefly introduced competing methodologies that were in section 3 used to estimate time-varying parameters and compare the results.

Section 3 introduced three parameters that were by the original research suspected to be time-varying. Two of the parameters changed during recent economic recession and one of the parameter change is associated with transformation period in the beginning of the data sample. The changes of the parameter captured by the two different methods are very similar.

Also, recursive impulse response analysis results show that the changes in the behavior of the economic system are similar when simulated with parameters obtained by filtration by Unscented Particle Filter and by Bayesian estimation.

To conclude, filtration with Unscented Particle Filter provides very interesting results which in many aspects correspond to the results of recursive analysis. The results obtained by Unscented Particle Filter are very promising and further research with this tool will no doubt be fruitful.

Acknowledgements

This work is supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/0780/2011. This work is supported by the Research Centre for the Competitiveness of the Czech Economy. The access to the MetaCentrum computing facilities provided under the programme "Projects of Large Infrastructure for Research, Development, and Innovations" LM2010005 funded by the Ministry of Education, Youth, and Sports of the Czech Republic is highly appreciated.

References

[1] Canova, F. What Explains The Great Moderation in the U.S.? A Structural Analysis. *Journal of the European Economic Association 7*, 4 (06 2009), 697–721.

[2] Clarida, R., Galí, J., and Gertler, M. Monetary Policy Rules And Macroeconomic Stability: Evidence And Some Theory. *The Quarterly Journal of Economics 115*, 1 (February 2000), 147–180.

[3] Čapek, J. Changes in the behavior of economies in a DSGE model in the course of time. In *Mathematical Methods in Economics 2011* (Janská Dolina, Slovakia, September 6 9 2011), V. S. Martin Dlouhý, Ed., Professional Publishing, pp. 94–100. ISBN 978-80-7431-058-4.

[4] Čapek, J. DSGE Model Sensitivity to Current Recession. In *Mathematical Methods in Economics* (2010), University of South Bohemia, Faculty of Economics, pp. 90–97. ISBN 978-80-7394-218-2.

[5] Fernández-Villaverde, J. and Rubio-Ramírez, J. F. How Structural Are Structural Parameters? In *NBER Macroeconomics Annual 2007, Volume 22*, NBER Chapters. National Bureau of Economic Research, Inc, march 2008, pp. 83–137.

[6] Lišťinský, Z. Time-varying Parameters in a Small Open Economy Model of the Czech Republic *mimeo* (2012)

[7] Lubik, T. and Schorfheide, F. A Bayesian Look at the New Open Economy Macroeconomics. In *NBER Macroeconomics Annual 2005, Volume 20*, NBER Chapters. National Bureau of Economic Research, Inc, May 2006, pp. 313–382.

[8] Tonner, J., Polanský, J., and Vašíček, O. Parameter drifting in a DSGE model estimated on czech data. *Czech Journal of Economics and Finance (Finance a úvěr) 61*, 5 (November 2011), 510–524.

Note on Optimal Paths for Non-Motorized Transport on the Network

Anna Černá[1], Jan Černý[2]

Abstract. The paper presents decision problems concerning the design of routes for non-motorized leisure transport, e.g. cycling or hiking, on the given network. We assume that the network is represented by a non-oriented graph. Three figures are given for each edge of the graph: construction cost, transit time and measure of attractiveness to users. Moreover, the attractiveness measure is defined for each vertex of the graph as well. The problem is to find a path with the duration not exceeding the given limit, maximizing the attractiveness and minimizing the construction cost. The paper describes several alternatives of solution of this two-objective problem.

Keywords: subnetwork, optimal, passenger transport, accessibility, cost.

JEL Classification: C65, R42
AMS Classification: 05C35, 90B06, 90B10

1 Introduction

The main purpose of the paper is to propose mathematical models and methods to support decision making on the choice of economically and 'culturally' optimal route for leisure cycling or hiking (but not both) between two network nodes.

These topics belong to the network economy theory. This scientific field contains two main directions: The first one studies networks of mutually interacting institutions and enterprises, as presented e.g. in the interesting books [5] and [8]. The second direction seeks for economically optimal telecommunication and transport network, see e.g. [14] and [11].

Our problem falls in the family of network reduction problems, namely the ones looking for a subnetwork of the given network. One can mention papers studying similar problems. The paper [6] looks for the maximum planar subgraph of the given graph G or [4] and [9] look for a k-edge connected spanning subgraph which minimizes the edge costs for the given $k > 0$ and similar can be found in [7] and [2] .Minimization of the costs of edges are dealt in [13] or [1] . Other papers seek for maximum cost subgraphs, see. [15] and [1].

As one can see from this survey, no paper from the abovementioned ones deals with non-motorized transport, neither serving for the commuting to work or school, nor for leisure. Therefore, the main contribution of this paper is the formulation of typical problems of leisure cycling and hiking routes design and the presentation of methods solving these problems.

The basic difference between cycle or hiking path for the commuting and the one for leisure ride/walk is that for the trip to work the shortest path is chosen, while for leisure the most attractive route of given yardage (or time) limit is preferred. Note that we use the term 'path' in the sense of 'simple path', which does not pass any intermediate node (=vertex) twice. We use the term 'route' when such multiple passing through vertices or edges is allowed. Even though we shall look for leisure routes, leisure (simple) paths are not a priori excluded.

Assume that there are two terminals of rail or bus transport v_A and v_B. Moreover, suppose that these terminals belong to the vertex set V of the given 'candidate' undirected graph $G = (V, E, a, c, t)$ for either leisure cycling, or hiking, but not both, where:

E is the edge set, each edge $e \in E$ represents a 'candidate' of transformation into the state suitable for leisure hiking or cycling traffic,

a is a non-negative function on the set $E \cup V$, $a(v)$ and $a(e)$ expressing a 'leisure attractivenes' of visiting the vertex $v \in V$ or the edge $e \in E$ respectively, specially $a(v_A) = a(v_B) = 0$,

[1] University of Economics in Prague, Faculty of Management, Jarošovská Street 1117/II, 37701 Jindřichův Hradec, cerna@fm.vse.cz.
[2] University of Economics in Prague, Faculty of Management, Jarošovská Street 1117/II, 37701 Jindřichův Hradec, cerny@fm.vse.cz.

c is a positive function on the set E, $c(e)$ expressing the cost of transformation the edge $e \in E$ into the state suitable for leisure hiking or cycling traffic.

t is a positive function on the set E, $c(e)$ expressing the passing time through the edge $e \in E$ for leisure hiking or cycling.

2 Optimization Problems

Problem P1: Let $G = (V, E, a, c, t)$ be the given graph, defined as above. Let $v_A \in V$, $v_B \in V$ be given vertices of G. Let $\tau > 0$ be a time limit for the duration of the trip. The problem is to find a path $p = (v_A = v_0, v_1, ..., v_n = v_B)$ such that

P1.1: $t(p) = t(v_0, v_1) + t(v_0, v_1) + ... + t(v_{n-1}, v_n) \leq \tau$

P1.2: $c(p) = c(v_0, v_1) + c(v_0, v_1) + ... + c(v_{n-1}, v_n) \rightarrow \min$

P1.3: $a(p) = a(v_0, v_1) + a(v_1) + a(v_0, v_1) + a(v_2) + ... + a(v_{n-1}) + a(v_{n-1}, v_n) \rightarrow \max$

Problem P2: Let $G = (V, E, a, c, t)$ be the given graph, defined as above. Let $A \in V$, $B \in V$ be given vertices of G. Let $\tau > 0$ be a time limit for the duration of the trip and let $\varphi \in (0, 1)$ be an attractiveness reduction factor. The problem is to find a route $r = (v_A = v_0, v_1, ..., v_n = v_B)$ such that $v_{i-1} \neq v_i$ for $i = 1, ..., n$ and

P2.1: $t(r) = t(v_0, v_1) + t(v_0, v_1) + ... + t(v_{n-1}, v_n) \leq \tau$.

P2.2: $c(r) = c(v_0, v_1) + c(v_0, v_1) + ... + c(v_{n-1}, v_n) \rightarrow \min$.

P2.3: $a(r) = a(v_0, v_1) + \varphi(v_1)a(v_1) + \varphi(v_1)(v_1, v_2)a(v_0, v_1) + \varphi(v_2)a(v_2) + ... + \varphi(v_{n-1})a(v_{n-1}) + \varphi(v_{n-1}, v_n)a(v_{n-1}, v_n) \rightarrow \max$.

There $\alpha(v_i, v_{i+1}) = 1$ if the route r passes through the edge (v_i, v_{i+1}) for the first time and $\alpha(v_i, v_{i+1}) = \alpha$ otherwise, and similarly $\alpha(v_i) = 1$ if the route r passes through the vertex v_i for the first time and $\alpha(v_i) = \alpha$ otherwise.

Although the problem P1 is a particular case of the problem P2, we shall deal with both since they possess different methods of solution.

Note: We use the term 'path' when no vertex is contained more than once in the sequence. If the multiple occurrence of vertices is allowed in the sequence, then we use the term 'route'.

We see that both problems are bicriterial, where the first criterion expresses economic aspects, the second one shows the leisure attractiveness. Theoretically, all known approaches of multicriterial analysis can be applied here but the current Czech situation leads to the applications where limited funding plays the dominant role. Therefore, we suppose a cost limit γ is given and both requirements P1.2 and P2.2 are reformulated to the form:

P1.2'≡ P2.2': $c(r) = c(v_0, v_1) + c(v_0, v_1) + ... + c(v_{n-1}, v_n) \leq \gamma$ ($r = p$ in the case of P1.2')

and then we shall speak about the problems P1' or P2' respectively.

The attractiveness function a is of a combined type, i.e. it is defined as for the edges as for the vertices of the graph. This may cause complication in the solution of the problems. Therefore, we define a new graph with extended vertex and edge sets $G'' = (V'', E'', a, c, t)$, where

$$V'' = (V \cup \{v': v \in V, a(v) > 0\} \cup \{v'': v \in V, a(v) > 0\}) - \{v \in V: a(v) > 0\} \tag{1}$$

$$\begin{aligned} E'' = {} & \{(v', v''): v' \in V''\} \cup \{(v, w): v \in V'', w \in V'', (v, w) \in E\} \cup \\ & \cup \{(v, w'): v \in V'', w' \in V'', (v, w) \in E\} \cup \{(v, w''): v \in V'', w'' \in V'', (v, w) \in E\} \cup \\ & \cup \{(v', w'): v' \in V'', w' \in V'', (v, w) \in E\} \cup \{(v', w''): v' \in V'', w'' \in V'', (v, w) \in E\} \cup \\ & \cup \{(v'', w''): v'' \in V'', wC \in V'', (v, w) \in E\} \end{aligned} \tag{2}$$

Moreover, if we define the symbol s as one of the possibilities: single quote (') or double quote (″) or empty space ( ), then we put

$$a(v', v'') = a(v) \text{ for each } v' \in V'' \text{ and } a(v^s, w^s) = a(v, w) \text{ for each } (v^s, w^s) \in E'' \tag{3}$$

$$c(v', v'') = 0 \text{ for each } v' \in V'' \text{ and } c(v^s, w^s) = c(v, w) \text{ for each } (v^s, w^s) \in E'', w \neq v \tag{4}$$

$$t(v', v'') = 0 \text{ for each } v' \in V'' \text{ and } t(v^s, w^s) = t(v, w) \text{ for each } (v^s, w^s) \in E'', w \neq v \tag{5}$$

$$a(w) = 0 \text{ for each } w \in V'' \tag{6}$$

Let $r = (v_A = v_0, v_1, ..., v_n = v_B)$ be s route on the graph G, such that $v_{i-1} \neq v_i$ for $i = 1, ..., n$. Then we define the image $r'' = f(r)$ on the graph G'' by replacing of each vertex v_i, having the property $a(v_i) > 0$, by a pair v_i', v_i''.

Lemma 1: f is one-to-one mapping of paths from the graph G onto the ones from G'' preserving values of the functions a, c, and t.

Proof: It is obvious that f is one-to-one mapping. The rest of the proof follows from (3), (4), (5) and (6).

Corollary: Since a path is a particular case of a route, we can solve problems P1' and P2' on the graph G'' instead of G.

Assumption: Throughout the rest of the text, we will assume that the problem P1' is solvable, i e. that there exists a path $p = (v_A = v_0, v_1, ..., v_n = v_B)$ such that $c(p) \leq \gamma$, $t(p) \leq t$.

3 Solution of the Problems P1′ and P2′

In accordance with the Corollary of the Lemma 1 we shall solve the **problem P1′** on the graph G'' where the attractiveness $a > 0$ is assigned only to edges. We shall use an **exact method** of the "Depth-First-Search" type.

The vertex v_A represents the root of the solution tree. All adjacent vertices v to v_A in G'' represent the first level vertices. All adjacent vertices w to all vertices v of the 1st level G'' represent the 2nd level vertices etc.

The search starts in the root v_A with the record $a_{rec} = 0$ and explores each branch as deep as possible. The path p from the root to the current vertex w in the solution tree represents a path in the graph G''. Backtracking is applied when at least one from the following situations occurs:

S1. w is in the path for the second time,

S2. $c(p) > \gamma$,

S3. $t(p) > t$,

S4. $w = v_B$; in that case if $a(p) > a_{rec}$ then the value a_{rec} is increased to $a(p)$.

After having finished the search, the last p is optimal.

Remark: This procedure can serve as a heuristics as well if one stops it before completing the search.

Solution of the **problem P2′** is almost the same as in the case of P1′. Only the first backtracking situation S1 is omitted and $a(r)$ may contain attractiveness of 2nd and further passing through the same edge.

4 Conclusion

In this paper, we have formulated two basic problems concerning the design of leisure routes between two points v_A and v_B for non-motorized tourists, i.e. for hikers or for cyclists (but not for both in the same problem). The first problem concerns simple paths, not passing twice the same point. The second problem deals with routes which are allowed to pass twice or more times through any point.

The path or the route are designed on the given 'candidate' network, where each edge e is assigned three non-negative numbers: $a(e)$ – the attractiveness of passing through e for tourists, $c(e)$ – the construction cost of e and $t(e)$ the passing time through e for a hiker or cyclist. Moreover, the attractiveness $a(v)$ is given for each vertex v. The first problem looks for a (simple) path p connecting the two given vertices v_A and v_B such that the duration $t(p)$ of the passing through the path p does not exceed the given limit t, the total construction cost $c(p)$ of the path p does not exceed the given limit γ and the total attractiveness of p is maximum. The second problem does the same for a general route r which is allowed to pass any vertex or any edge more than once.

A 'Depth-First-Search'-type exact method is presented for both problems and it is noted that the method can be used also as a heuristics if it is stopped before having examined the whole tree of solutions.

The future research can be expected in two main directions. The first one is methodological. As concerns the problem P1, we hope to find a LP formulation of the problem and, moreover, it is likely that there will be

found a dual type of heuristics. Maybe, an inspiration can be found in [3], since the problems studied there are similar to P1. Similar expectations can be formulated concerning the problem P2, since it is a bit related to the problem of [10] and [12].

The second direction of the future research will be oriented on the practical experience in applying models and methods described above.

Acknowledgements

The support of the Czech Science Foundation # 402/12/2147 'Economically Optimal Processes on Networks' is gratefully acknowledged.

References

[1] Amini, O., Peleg, D., Pérennes, S., Sau, I., Saurabh, S.: Degree-Constrained Subgraph Problems: Hardness and Approximation Results. In: *WAOA 2008, LNCS 5426* (Bampis, E. and Skutella, M. eds.), Springer-Verlag Berlin Heidelberg 2009, 29–42.

[2] Carr, R., Ravi, R.: A New Bound for the 2-Edge Connected Subgraph Problem. In: *IPCO VI, LNCS 1412* (Bixby, R. E., Boyd, E. A., and Rios-Mercado, R. Z. eds.), Springer-Verlag Berlin Heidelberg 1998, 112-125.

[3] Černá, A., Černý, J., Přibyl, V.: Bus Route Design in Small Demand Areas. *Transport*, 26, 3 (2011), pp. 248-254.

[4] Cheriyan, J., Thurimella, R.: Approximating Minimum-Size k-Connected Spanning Subgraphs via Matching, *SIAM J. Comput.*, 30, 2 (2000), 528-560.

[5] De Man, A.P.: *The Network Economy: Strategy, Structure and Management*. Edward Elgar, Cheltenham, GB, 2004.

[6] Eades, P., Foulds, L. R., Giffin, J. W.: An efficient Heuristic for Identifying a Maximum Weight Planar Subgraph. *Combinatorial Mathematics IX, Lecture Notes in Mathematics No. 952*, Springer-Verlag, Berlin 1982.

[7] Fernandes, C.G.: A Better Approximation Ratio for the Minimum Size k-Edge-Connected Spanning Subgraph Problem, *J. Algorithms*, 28, 1 (1998), 105-124.

[8] Fiala, P.: *Network Economics* (in Czech) Professional Publishing, Praha 2008

[9] Gabow, H.N., Goemans, M.X., Tardos, E., Williamson, D.P.: Approximating the Smallest k-Edge Connected Spanning Subgraph by LP-Rounding. In: *SODA '05 Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics Philadelphia, PA, USA, 2005, 562-571.

[10] Jerby, S., Ceder, A. Optimal routing design for shuttle bus service. *Transportation Research Record*, No.1971, (2006), pp.14-22

[11] Jun Xiao et al.: Comprehensive method on evaluation of distribution network planning. In: *Proc. DRPT 2008, Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, Nanjuing (China), April 2008, 1249-1254.

[12] Kuo, Y., Luo. C.M., Wang, C.C., Optimization of circular-type routes for a shuttle bus service, *Transportation Research Record 90th Annual Meeting,* Washington DC.

[13] Safari, M.A., Salavatipour, M.R.: A Constant Factor Approximation for Minimum λ-Edge-Connected k-Subgraph with Metric Costs. In: *APPROX and RANDOM 2008, LNCS 5171* (Goel A. et al. eds.), Springer-Verlag Berlin Heidelberg 2008, 233–246.

[14] Shimoda, Y.: Heat source network and heat transport. In: *District Heating & Cooling* Vol. 41 (2000), No. 3, pp.3-5.

[15] Suzuki, A. Tokuyama, T.: Dense subgraph problem revisited. *Proc. Joint Workshop "New Horizons in Computing" and "Statistical Mechanical Approach to Probabilistic Information Processing"* ,18-21 July, 2005, Sendai, Japan.

A note on the choice of a sample of firms for reliable estimation of sector returns to scale

Michal Černý[1]

Abstract. A sector is defined as a family of firms sharing the same Cobb-Douglas production function. Our aim is to estimate the Cobb-Douglas-based returns to scale of the sector. Limited resources allow us to collect data (stock of production factors and production) from a limited number of firms only. We address the question how the sample of firms, used then for estimation of the sector returns to scale, should be selected to achieve a "good" estimate of the returns to scale. (The estimate is "good" if it has low variance.) We propose a three-step procedure for the sample selection problem, adopting a method from the theory of c-optimal experimental designs. We consider both homoscedastic and heteroscedastic models. We illustrate the approach by examples.

Keywords: sample selection, Cobb-Douglas function, returns to scale, c-optimal design

JEL classification: C81
AMS classification: 62K05, 91B38, 91G70

1 Introduction, definitions and assumptions

Let $\Phi_1, \ldots, \Phi_n$ denote production factors. A *firm* is a $(n+1)$-tuple of nonnegative real numbers

$$(y^*, \varphi_1, \ldots, \varphi_n), \tag{1}$$

where y^* denotes the level of the firm's output and φ_i denotes the stock of i-th production factor available to the firm.

A *sector* $\mathcal{S}$ is the set

$$\mathcal{S} = \{F_1, \ldots, F_N\},$$

where $F_1, \ldots, F_N$ are firms. We also use the notation

$$F_j = (y_j^*, \varphi_{1j}, \ldots, \varphi_{nj}). \tag{2}$$

We assume that all the firms of the sector $\mathcal{S}$ share a common Cobb-Douglas production function of the form

$$\ln y_j = \beta_0 + \sum_{i=1}^{n} \beta_i \ln \varphi_{ij} + \varepsilon_j, \qquad j = 1, \ldots, N, \tag{3}$$

where ε_j are independent $N(0, \sigma^2)$ error terms. In (2) we assume that the value y_j^* is the observed realization of the random variable y_j.

Returns to scale of the sector $\mathcal{S}$ is the number $r := \sum_{i=1}^{n} \beta_i$. Recall that the returns to scale are

$$\left.\begin{array}{r}\text{constant}\\ \text{increasing}\\ \text{decreasing}\end{array}\right\} \text{ iff } \left\{\begin{array}{l} r = 1,\\ r > 1,\\ r < 1.\end{array}\right.$$

[1]University of Economics Prague, Department of Econometrics, Winston Churchill Square 4, 13067 Prague, Czech Republic, cernym@vse.cz

1.1 The problem

Our aim is to measure the number r of the sector $\mathcal{S}$. Of course, due to the presence of the error terms ε_j, we can never measure r exactly. Therefore we are interested in an *estimate* $\widehat{r}$ of r. We shall use the standard estimator $\widehat{r} = \sum_{i=1}^{n} \widehat{\beta}_i$, where $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \ldots, \widehat{\beta}_n)^{\mathrm{T}}$ is the standard OLS estimator of (3). Then we can, for example, test the null hypothesis

$$r = 1 \tag{4}$$

using the standard t-test or F-test.

Assume that N, the size of the sector, is large. In order to obtain as precise estimates of r as possible, it is desirable to collect data (2) for all firms $F_1, \ldots, F_N$. However, this process is usually costly. Usually only limited resources are available to us; with these resources we are able to collect data from a limited number of firms only. We have arrived at the main question of the paper: *assume that we are able to collect data from only $m \ll N$ firms. Which firms from the sector $\mathcal{S}$ should be included in the selected sample $\mathcal{S}'$ (of cardinality m) in order the value $\widehat{r}$, estimated from the sample $\mathcal{S}'$, be as precise as possible?*

The relevance of the question is motivated by the following example.

1.2 Example

Assume that $n = 2$ and $\Phi_1 = \textit{labor}$ and $\Phi_2 = \textit{capital stock}$. Assume that the sector $\mathcal{S}$ of $N = 12$ firms is governed by the model (3) with

$$\beta_0 = 0, \quad \beta_1 = 0.5, \quad \beta_2 = 0.6, \quad \sigma = 0.1.$$

Then $r > 1$ and the returns to scale of the sector $\mathcal{S}$ are increasing.

Assume that our resources allow us to gather data from $m = 6$ firms only. We would like to choose the sample of 6 firms in the way that $\mathrm{se}(\widehat{\beta}_1 + \widehat{\beta}_2)$ is minimal, where "se" stands for standard error. In that case, r is estimated with the best possible precision. This is important since the standard error of $\widehat{r}$ being low, the t-test for the hypothesis $r = 1$ is strong. (Recall that the test statistic is of the form $\frac{\widehat{r}-1}{\widehat{\mathrm{se}}(\widehat{r})}$.)

We can write the model (3) in the usual form $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^{\mathrm{T}}$. With this notation we have

$$\mathrm{se}(\widehat{\beta}_1 + \widehat{\beta}_2) = \sigma \cdot \sqrt{\boldsymbol{c}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{c}},$$

where $\boldsymbol{c} = (0, 1, 1)^{\mathrm{T}}$.

We have $\binom{12}{6} = 924$ possibilities for the choice of the sample $\mathcal{S}'$ of 6 firms out of 12 total; denote the choices as $\mathcal{S}'_1, \ldots, \mathcal{S}'_{924}$. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{924}$ denote the corresponding $\boldsymbol{X}$-matrices. Define

$$\tau_i := \sigma \cdot \sqrt{\boldsymbol{c}^{\mathrm{T}}(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{X}_i)^{-1}\boldsymbol{c}}, \quad i = 1, \ldots, 924.$$

Let the choices $\mathcal{S}'_1, \ldots, \mathcal{S}'_{924}$ be ordered in the way that $\tau_1 \leq \cdots \leq \tau_{924}$. Figure 1 shows values of τ_i against i. The best possible choice is

$$\mathcal{S}'_1 = \{F_1, F_2, F_3, F_6, F_8, F_9\} \text{ with } \tau_1 = 0.0435, \tag{5}$$

while the worst possible choice is

$$\mathcal{S}'_{924} = \{F_4, F_5, F_7, F_8, F_{11}, F_{12}\} \text{ with } \tau_{924} = 0.2064. \tag{6}$$

In the case (6), t-test for the null hypothesis (4) will probably not reject, though the hypothesis is not true. Hence, with the choice $\mathcal{S}'_{924}$ we can arrive at an incorrect conclusion that returns to scale are constant. On the other hand, if we choose the sample $\mathcal{S}'_1$, we have a much higher chance that the t-test will reject, which is a correct conclusion. In general: the better value τ_i, the stronger the t-test is. And, if we choose the sample of firms "in the best possible way" and the t-test does not reject, we have a strong evidence that $r = 1$ indeed.

This example shows that before we start collecting data, *it is reasonable to ask which firms of the sector $\mathcal{S}$ are likely to contribute to the precision of the estimator of $\widehat{r}$ more than others.*

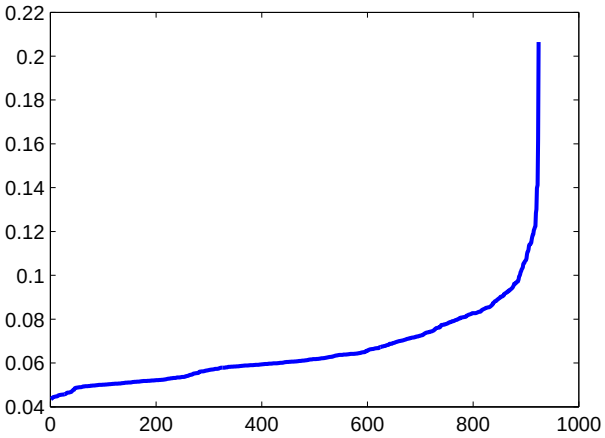


Figure 1 Sequence $\tau_1, \ldots, \tau_{924}$.

2 Our approach

The question leads us to the theory of optimum experimental designs. Indeed, the sample which minimizes the variance of $\widehat{r}$ can be seen as a case of $\boldsymbol{c}$-optimal design: our aim is minimization of $\mathrm{se}(\boldsymbol{c}^{\mathrm{T}}\widehat{\boldsymbol{\beta}})$, where $\boldsymbol{c} = (0, 1, \ldots, 1)^{\mathrm{T}}$ and $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_n)^{\mathrm{T}}$.

The problem is that we know nothing about the sector $\mathcal{S}$ in advance. We adopt the assumption that we are able to gather information on *representants* of the sector $\mathcal{S}$. Each representant should represent a group of firms in the sector $\mathcal{S}$ with similar stock of production factors. (Said more precisely, a representant R should be either a real or fictitious firm such that it is reasonable to expect that in the sector $\mathcal{S}$ there are enough real firms with the stock of production factors similar to R.) Then we restrict ourselves to the representants.

We find an optimal design over the representants; this will give us guidance from which groups of firms it should be suitable to collect final data.

We illustrate the approach by example. Let φ_{1j} denote the capital stock of j-th representant and let φ_{2j} denote the labor stock of j-th representant. Assume that we know that the sector $\mathcal{S}$ contains the following groups with the following representants:

group	type	representant
group 1	small capital-intensive firms	$R_1 = (\varphi_{11} = 5,\ \varphi_{21} = 1)$
group 2	small labor-intensive firms	$R_2 = (\varphi_{12} = 1,\ \varphi_{22} = 5)$
group 3	medium capital-intensive firms	$R_3 = (\varphi_{13} = 20,\ \varphi_{23} = 10)$
group 4	medium labor-intensive firms	$R_4 = (\varphi_{14} = 15,\ \varphi_{24} = 22)$
group 5	large capital-intensive firms	$R_5 = (\varphi_{15} = 35,\ \varphi_{25} = 20)$
group 6	large labor-intensive firms	$R_6 = (\varphi_{16} = 20,\ \varphi_{26} = 42)$

(7)

In our example we will write

$$\mathcal{X} := \left\{ \begin{pmatrix} 1 \\ \ln \varphi_{11} \\ \ln \varphi_{21} \end{pmatrix}, \ldots, \begin{pmatrix} 1 \\ \ln \varphi_{16} \\ \ln \varphi_{26} \end{pmatrix} \right\}. \tag{8}$$

The meaning of this set will be explained in the next section.

2.1 Some notions from the theory of c-optimal designs

In the theory of experimental design, the set $\mathcal{X}$ is usually referred to as *experimental domain*. Its interpretation is as follows. Assume the linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{9}$$

with independent disturbances $\boldsymbol{\varepsilon}$, which are homoscedastic with variance σ^2. We are given a nonzero vector $\boldsymbol{c}$ of parameters and our aim is to select the rows of $\boldsymbol{X}$ in the way that $\mathrm{se}(\boldsymbol{c}^{\mathrm{T}}\widehat{\boldsymbol{\beta}})$ is minimal. We are restricted by the fact that each row $\boldsymbol{x}^{\mathrm{T}}$ of $\boldsymbol{X}$ must fulfill $\boldsymbol{x} \in \mathcal{X}$. Said otherwise, we can make measurements only in the points from the experimental domain $\mathcal{X}$ and our aim is to select those points which minimize the variance of $\boldsymbol{c}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}$.

Assume that $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ and that we have the regression model (9) with ν observations, where the matrix $\boldsymbol{X}$ is of the form

$$\boldsymbol{X} = (\underbrace{\boldsymbol{x}_1, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_1}_{\nu\xi_1 \text{ times}}; \ \underbrace{\boldsymbol{x}_2, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_2}_{\nu\xi_2 \text{ times}} \ \cdots \ ; \underbrace{\boldsymbol{x}_M, \boldsymbol{x}_M, \ldots, \boldsymbol{x}_M}_{\nu\xi_M \text{ times}})^{\mathrm{T}}. \tag{10}$$

The vector $\boldsymbol{\xi} := (\xi_1, \ldots, \xi_M)^{\mathrm{T}}$ is called *design* — it simply says that we are making $100\xi_1\%$ observations in the point $\boldsymbol{x}_1$, $100\xi_2\%$ observations in the point $\boldsymbol{x}_2$ etc.

We can define the number $\mathrm{var}_{\boldsymbol{c}}(\boldsymbol{\xi})$, called $\boldsymbol{c}$*-variance* of the design $\boldsymbol{\xi}$, implicitly using the equation

$$\mathrm{var}(\boldsymbol{c}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}) = \frac{\sigma^2}{\nu} \cdot \mathrm{var}_{\boldsymbol{c}}(\boldsymbol{\xi}),$$

where $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}$ with $\boldsymbol{X}$ given by (10). (Here, $^{-1}$ might stand for the matrix pseudoinverse.) It is easy to see that the number $\mathrm{var}_{\boldsymbol{c}}(\boldsymbol{\xi})$ does depend on the design $\boldsymbol{\xi}$, *but it depends neither on* σ^2 *nor on the number of observations* ν. Hence it is a good measure of the contribution of the design $\boldsymbol{\xi}$ to the total variance of the estimator $\boldsymbol{c}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}$.

All designs form the simplex $\Sigma := \{\boldsymbol{\xi} : \boldsymbol{\xi} \geq \boldsymbol{0},\ \boldsymbol{1}^{\mathrm{T}}\boldsymbol{\xi} = 1\}$. Our task is to find the design with minimal $\boldsymbol{c}$-variance. Thus we are to solve the optimization problem

$$\min\{\mathrm{var}_{\boldsymbol{c}}(\boldsymbol{\xi}) : \boldsymbol{\xi} \in \Sigma\}.$$

Its solution is called $\boldsymbol{c}$*-optimal design.*

Definition 1. The **Elfving set** is the set $\mathcal{E} := \mathrm{convexhull}(\mathcal{X} \cup -\mathcal{X})$, where $-\mathcal{X} = \{-\boldsymbol{x} : \boldsymbol{x} \in \mathcal{X}\}$. □

The following theorem, called Elfving's Theorem (see [4]), is a fundamental result in the theory of $\boldsymbol{c}$-optimal designs.

Theorem 1. *Let* $\boldsymbol{c}$ *be a nonzero vector and let* $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$. *Let* $\omega^* = \max\{\omega \in \mathbb{R} : \omega \cdot \boldsymbol{c} \in \mathcal{E}\}$ *and* $\boldsymbol{x}^* = \omega^*\boldsymbol{c}$. *Let* $u_1, \ldots, u_M$ *and* $v_1, \ldots, v_M$ *be nonnegative numbers such that*

$$\boldsymbol{x}^* = \sum_{i=1}^{M} u_i\boldsymbol{x}_i - \sum_{i=1}^{M} v_i\boldsymbol{x}_i$$

and

$$\sum_{i=1}^{M}(u_i + v_i) = 1.$$

Then $(u_1 + v_1, \ldots, u_M + v_M)^{\mathrm{T}}$ *is the* $\boldsymbol{c}$*-optimal design over* $\mathcal{X}$. □

In other words, if we write the point $\boldsymbol{x}^*$ as a convex combination of the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M, -\boldsymbol{x}_1, \ldots, -\boldsymbol{x}_M$, then the coefficients of the convex combination determine the $\boldsymbol{c}$-optimal design.

Harman and Jurík [5] observed that Elfving's Theorem leads to a linear programming problem.

Theorem 2. *Let* $\boldsymbol{\Xi} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M)$. *Let* $\boldsymbol{u}^*, \boldsymbol{v}^*, \omega^*$ *be the solution of the linear program*

$$\max\{\omega \in \mathbb{R} :\ \boldsymbol{\Xi}(\boldsymbol{u} - \boldsymbol{v}) = \omega \cdot \boldsymbol{c},\quad \boldsymbol{1}^{\mathrm{T}}(\boldsymbol{u} + \boldsymbol{v}) = 1,\quad \boldsymbol{u} \geq \boldsymbol{0},\quad \boldsymbol{v} \geq \boldsymbol{0}\}. \tag{11}$$

Then $\boldsymbol{\xi} := \boldsymbol{u}^* + \boldsymbol{v}^*$ *is the* $\boldsymbol{c}$*-optimal design.* □

More on the theory of optimal designs can be found in [2], [6], [7]. Computational issues are dealt with in [1], [3].

2.2 The example continued

We now apply Elfving's Theorem to the "experimental domain" $\mathcal{X}$ given by (8). (The form of the model (3) shows why the logarithms are present in (8).) We set

$$\boldsymbol{\Xi} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ \ln 5 & \ln 1 & \ln 20 & \ln 15 & \ln 35 & \ln 20 \\ \ln 1 & \ln 5 & \ln 10 & \ln 22 & \ln 20 & \ln 42 \end{pmatrix}$$

and

$$\boldsymbol{c} = (0, 1, 1)^{\mathrm{T}}.$$

Solving the linear program (11) we get the optimal design $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_6)^{\mathrm{T}}$ with

$$\xi_1 = 0.13, \quad \xi_2 = 0.37, \quad \xi_3 = \xi_4 = \xi_5 = 0, \quad \xi_6 = 0.5. \tag{12}$$

This shows that we should compose the sample as follows:

- 13% of the observations should be collected from the group represented by the representant R_1,
- 37% of the observations should be collected from the group represented by the representant R_2,
- 50% of the observations should be collected from the group represented by the representant R_6.

If our budget is limited to, say, $m = 100$ firms, then it is reasonable to collect data from

- 13 small capital-intensive firms,
- 37 small labor-intensive firms and
- 50 large labor-intensive firms.

2.3 The heteroscedastic case

In the analysis of production functions it is often reasonable to assume heteroscedasticity. Let us consider an example with a heteroscedasticity model where the standard error of disturbances is proportional to $\sqrt{\varphi_{1j}\varphi_{2j}}$ (again, φ_{1j} denotes the capital stock of j-th firm and φ_{2j} denotes the labor stock of j-th firm). Then we can write the model (3) in the form

$$\ln y_j = \beta_0 + \beta_1 \ln \varphi_{1j} + \beta_2 \ln \varphi_{2j} + \delta_j \sqrt{\varphi_{1j}\varphi_{2j}},$$

where δ_j are independent and homoscedastic. A simple transformation yields

$$\frac{\ln y_j}{\sqrt{\varphi_{1j}\varphi_{2j}}} = \beta_0 \cdot \frac{1}{\sqrt{\varphi_{1j}\varphi_{2j}}} + \beta_1 \cdot \frac{\ln \varphi_{1j}}{\sqrt{\varphi_{1j}\varphi_{2j}}} + \beta_2 \cdot \frac{\ln \varphi_{2j}}{\sqrt{\varphi_{1j}\varphi_{2j}}} + \delta_j,$$

which is a homoscedastic model, and we can apply Elfving's Theorem. Using again the representants from (7), we set

$$\boldsymbol{\Xi} = \begin{pmatrix} \frac{1}{\sqrt{5\cdot 1}} & \frac{1}{\sqrt{1\cdot 5}} & \frac{1}{\sqrt{20\cdot 10}} & \frac{1}{\sqrt{15\cdot 22}} & \frac{1}{\sqrt{35\cdot 20}} & \frac{1}{\sqrt{20\cdot 42}} \\ \frac{\ln 5}{\sqrt{5\cdot 1}} & \frac{\ln 1}{\sqrt{1\cdot 5}} & \frac{\ln 20}{\sqrt{20\cdot 10}} & \frac{\ln 15}{\sqrt{15\cdot 22}} & \frac{\ln 35}{\sqrt{35\cdot 20}} & \frac{\ln 20}{\sqrt{20\cdot 42}} \\ \frac{\ln 1}{\sqrt{5\cdot 1}} & \frac{\ln 5}{\sqrt{1\cdot 5}} & \frac{\ln 10}{\sqrt{20\cdot 10}} & \frac{\ln 22}{\sqrt{15\cdot 22}} & \frac{\ln 20}{\sqrt{35\cdot 20}} & \frac{\ln 42}{\sqrt{20\cdot 42}} \end{pmatrix}$$

and $\boldsymbol{c}^{\mathrm{T}} = (0, 1, 1)$. Solution of the linear program (11) yields

$$\xi_1 = 0.1, \quad \xi_2 = 0.04, \quad \xi_3 = 0.86, \quad \xi_4 = \xi_5 = \xi_6 = 0. \tag{13}$$

So, if we are restricted to $m = 100$ observations, it is reasonable to collect data from

- 10 small-sized capital intensive firms,
- 4 small-sized labor-intensive firms and
- 86 medium-sized capital-intensive firms.

3 Conclusion

The difference between (12) and (13) shows that the homoscedasticity/heteroscedasticity assumption is important. (This is not surprising.) We thus suggest that it could be reasonable to perform the analysis in three steps:

- **Step 1.** Make a rough screening of the sector $\mathcal{S}$ to
  - identify groups of firms and their representants,
  - determine whether heteroscedasticity is present, and if so, estimate a suitable model of heteroscedasticity.
- **Step 2.** Using the data from Step 1, apply the method of Section 2.2 (if heteroscedasticity is not present) or Section 2.3 (if heteroascedasticity is present): find the optimal design $\boldsymbol{\xi}$ using (11).
- **Step 3.** Choose firms according to the design $\boldsymbol{\xi}$.

Acknowledgements

The work was supported by the project P403/12/1947 of the Czech Science Foundation.

References

[1] Antoch, J., Černý, M. and Hladík, M.: On computational complexity of construction of c-optimal linear regression models over finite experimental domains. *Tatra Mountains Mathematical Publications* **5** (2012), 1–10.

[2] Atkinson, A., Donev, A. and Tobias, R.: *Optimum Experimental Designs with SAS*. Oxford University Press, Oxford, 2007.

[3] Černý, M. and Hladík, M.: Two complexity results on c-optimality in experimental design. *Computational Optimization and Applications* **51** (2012), 1397–1408.

[4] Elfving, G.: Optimum allocation in linear regression theory. *Annals of the Institute of Statistical Mathematics* **23** (1952), 255–262.

[5] Harman, R. and Jurík, T.: Computing c-optimal experimental designs using the simplex method of linear programming. *Computational Statistics and Data Analysis* **53** (2008), 247–254.

[6] Pázman, A.: *Foundations of Optimum Experimental Design*. Reidel Publishing Company, Dordrecht, 1986.

[7] Pukelsheim, F. and Rieder, S.: Efficient rounding in approximate designs. *Biometrika* **79** (1992), 763–770.

Identifiability issue in macroeconomic modelling

Ondřej Čížek[1]

Abstract. The effort of central banks to apply DSGE models in practice has increased the complexity of these models. Recently however, it has been pointed out that even medium scale DSGE models chronically suffer from the identification problem, which is a topic that has been traditionally mentioned in macroeconomic modeling only in connection with the simultaneous equations models. The problem of identifiability has, however, far more general nature and apart from other consequences complicates practical econometric estimation of the parameters. The aim of this paper is to investigate the identifiability of the formulated small macroeconomic model by means of which I will show that the problem with identifiability can arise not only as a consequence of the model being too large as is frequently argued nowadays, but as a result of the model being too small. The formulated model falls within the class of linear models with quadratic loss function and its parameters are estimated by the method of maximum likelihood. The approach to analyze identifiability issue is based on the Fisher information matrix.

Keywords: identification, maximum likelihood, Fisher matrix, Kalman filter, DSGE, inflation targeting, macroeconomics.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The goal of this paper is to present key results from the theory of identification and to illustrate how to apply them to the formulated macroeconomic model, which falls within the class of linear models with quadratic loss function. To analyze identifiability issue means to answer the question whether or not different parameter values generate different probabilistic distribution of the observed variables. The chosen approach answers this question without a reference to a particular data set because data generating process (and not a particular data set) is used in the analysis. The analysis of identifiability, however, serves largely for the econometric estimation of the parameters, which is also a topic of this paper. For this reason, data used to estimate parameters by maximum likelihood method are described in chapter 3 immediately after the description of the model equations in chapter 2. Section 4 puts the model in convenient state space form. Chapter 5 then describes the Kalman filter and the next section shows how to use it to compute likelihood function. The outcome from the Kalman filter is also used to compute the Fisher information matrix as the main tool to detect problems with identifiability, which is a topic of the Section 7. Final chapter 8 discusses the results from the analysis of identifiability when applied to the formulated model.

2 Model

The presented model is a modification of the Ball's [1] macroeconomic model very well known in the literature on inflation targeting. Because the model is aimed at analyzing stabilization policy, all the variables represent cyclical component of the original value. There are two main variables in the model representing real economic activity and prices, which are the rate of unemployment (u) and the inflation (π). The first one is modeled by its own lagged value and by the interest rate (r), which is assumed to be under full control of the central bank:

$$u_{t+1} = a \cdot u_t + b \cdot r_{t+1} + \eta_{t+1}, \tag{1}$$

whereas $a \in (0,1)$, $b > 0$ are parameters and η_{t+1} is random shock, which is assumed to be normally distributed with zero mean, constant variance $\mathrm{var}(\eta_t) = \sigma_\varepsilon^2$ and is also assumed to be independent at time as well as with other random shocks in the model.

The equation (1) is in line with Ball's formulation. I modified, however, the relation between real economic activity and inflation represented by the Phillips curve in order to fit the data for the Czech economy. Ball assumed so called accelerationist Phillips curve in the form that $\pi_{t+1} - \pi_t$ is a function of u_t. The name of this

[1] University of Economics, Department of Econometrics, nám. W. Churchilla 4, Prague 3, cizeko@vse.cz.

type of Phillips curve comes from the fact that inflation raises if unemployment is below its "natural" level. My formulation is based not on this historical experience from the United States, but on the empirical data from the Czech Republic, which suggested negative relation between π_{t+1} and $u_t - u_{t-1}$. This relation can be interpreted in such a way that inflation gets above its "normal" level whenever (monetary) authority is trying to decrease the rate of unemployment. The equation for the inflation is as follows:

$$\pi_{t+1} = -c \cdot \left(u_t - u_{t-1}\right) + \varepsilon_{t+1} \tag{2}$$

where $c > 0$ is parameter and ε_{t+1} is random shock fulfilling same assumptions as η_{t+1}.

Ball completes the specification of the model by assuming that interest rate is set according to the Taylor rule[2] and then is trying to find such values of the coefficients of this rule that minimize $\mathrm{var}\left(\pi_t\right) + \kappa \cdot \mathrm{var}\left(u_t\right)$, where $\kappa > 0$ is the weight that is set on stabilization of the unemployment. In this article, I will use slightly more general approach and will assume the monetary authority to minimize quadratic loss function of the form:

$$E_0\left[\sum_{t=0}^{T} \delta^t \cdot \left(\kappa \cdot u_t^{\,2} + \pi_t^{\,2}\right)\right] \rightarrow MIN \tag{3}$$

which is equivalent to the Ball's approach if the discount factor of the monetary authority $\delta \in (0,1]$ is set to equal one and the horizont for optimization T goes to infinity.

3 Data

The source of data for inflation and rate of unemployment is the Czech Statistical Office. The inflation was calculated from consumer price index (CPI) on a quarterly basis according to $inflat_{t+1} = 100 \cdot \left(CPI_{t+1} - CPI_t\right) / CPI_t$. Although CPI is available from 1997 to 2011, I decided to work with data from 1999 because during years 1997 and 1998 the inflation was unstable[3]. There was a significant seasonal pattern in the calculated inflation, which I removed by means of linear regression with dummy variables Q_{it}. To model the increase in DPH at the beginning of the year 1998 from five to nine percentage points and at the beginning of the year 2000 from nine to ten percentage points, I also included another dummy variable DPH_t with value equal to four in the first quarter of the year 1998, one in the first quarter of the year 2000 and zero in all other quarters. All the dummies were statistically significant except the dummy for the third quarter. The cyclical component of the inflation was therefore calculated as residuals from the following regression:

$$inflat_t = \beta_1 \cdot Q_{1t} + \beta_2 \cdot Q_{2t} + \beta_4 \cdot Q_{4t} + \beta_5 \cdot DPH_t + error_t\,. \tag{4}$$

Data for the rate of unemployment was already seasonally adjusted by the Czech Statistical Office. I used the data from the selective survey of the labor force. The cyclical component of the rate of unemployment was calculated simply as a deviation from the mean value.

4 State space form

Although there are no unobserved variables in the model, it will be useful for the purpose of generalization to write the model in a state space form. If I define the state vector $\mathbf{x_t} = \left(\pi_t \quad u_t \quad u_{t-1}\right)'$ and the vector of random errors denote as $\mathbf{u_{t+1}} = \left(\varepsilon_{t+1} \quad \eta_{t+1}\right)'$, then the transition equation can be written in a following manner:

$$\mathbf{x_{t+1}} = \mathbf{A} \cdot \mathbf{x_t} + \mathbf{b} \cdot r_{t+1} + \mathbf{C} \cdot \mathbf{u_{t+1}}\,, \tag{5}$$

[2] Taylor rule specifies interest rate is a linear function of inflation and unemployment.

[3] The instability was caused by huge disturbances in the foreign exchange market in May 1997, which led to the change in monetary strategy of the central bank to inflation targeting at the beginning of the year 1998.

where $\mathbf{A} = \begin{pmatrix} 0 & -c & c \\ 0 & a & 0 \\ 0 & 1 & 0 \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 0 \\ b \\ 0 \end{pmatrix}$, $\mathbf{C} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$.

If I denote the vector of observed variables as $\mathbf{z}_t = \begin{pmatrix} \pi_t & u_t \end{pmatrix}'$, the measurement equation is as follows:

$$\mathbf{z}_t = \mathbf{D} \cdot \mathbf{x}_t , \tag{6}$$

where $\mathbf{D} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$.

The quadratic loss function can be written in the following way:

$$\underset{0}{E}\left(\sum_{t=0}^{\infty} \mathbf{x}_t' \cdot \mathbf{K}_t \cdot \mathbf{x}_t \right) \rightarrow MIN , \tag{7}$$

where $\mathbf{K}_t = \delta^t \cdot diag\begin{pmatrix} 1 & \kappa & 0 \end{pmatrix}$ is the weighting matrix.

5 Optimising control variables and estimating unobserved state

In this chapter I will suppose that the values of the parameters are known and show the results telling us how to solve the problem of minimising the objective function (7) as well as the problem of estimation of the unobserved state vector by the Kalman filter. In the next chapter, it will be shown how to use these result in econometric estimation of the parameters.

The aim of the optimal control problem is to find trajectory of control variables r_t, $t = 0,...,T$ that minimize the loss function (7). Derivation of the solution to this problem can be found for example in Chow [2], chapters 8.3 and 8.4. The solution is as follows:

$$r_t = \mathbf{G}_t \cdot \mathbf{x}_{t-1|t-1} , \tag{8}$$

where the matrix $\mathbf{G}_t$ is calculated recursively backwards in time according to the scheme:

$$\mathbf{G}_t = -\left(\mathbf{b}'\mathbf{H}_t\mathbf{b} \right)^{-1} \mathbf{b}'\mathbf{H}_t\mathbf{A} , \tag{9}$$

$$\mathbf{H}_{t-1} = \mathbf{K}_{t-1} + \left(\mathbf{A} + \mathbf{b}\mathbf{G}_t \right)' \mathbf{H}_t \left(\mathbf{A} + \mathbf{b}\mathbf{G}_t \right) , \tag{10}$$

with the initial value $\mathbf{H}_T = \mathbf{K}_T$. The symbol $\mathbf{x}_{t-1|t-1}$ denotes optimal estimation of the unobserved state vector, which is given by the Kalman filter recursions of the form:[4]

$$\mathbf{x}_{t|t-1} = \left(\mathbf{A} + \mathbf{b}\mathbf{G}_t \right) \cdot \mathbf{x}_{t-1|t-1} , \tag{11}$$

$$\mathbf{x}_{t|t} = \mathbf{x}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{D}' \left(\mathbf{D}\mathbf{P}_{t|t-1}\mathbf{D}' \right)^{-1} \left(\mathbf{z}_t - \mathbf{D}\mathbf{x}_{t|t-1} \right) , \tag{12}$$

where $\mathbf{x}_{t|t-1} \equiv \underset{t-1}{E}\left(\mathbf{x}_t \right)$ and matrices $\mathbf{P}_{t|t-1} \equiv \underset{t-1}{E}\left(\mathbf{x}_t - \mathbf{x}_{t|t-1} \right)\left(\mathbf{x}_t - \mathbf{x}_{t|t-1} \right)'$ are calculated recursively as follows:

$$\mathbf{P}_{t+1|t} = \mathbf{A}\left[\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{D}' \left(\mathbf{D}\mathbf{P}_{t|t-1}\mathbf{D}' \right)^{-1} \mathbf{D}\mathbf{P}_{t|t-1} \right] \mathbf{A}' + \mathbf{C}\Sigma_{uu}\mathbf{C}' , \tag{13}$$

[4] Because no prior information about initial value of the state vector was available, the Kalman filter recursion was initialized by so called "diffuse prior", that is by setting $\mathbf{x}_{0|-1} = \mathbf{0}$, $\mathbf{P}_{0|-1} = \alpha \cdot \mathbf{I}$, where $\mathbf{I}$ is identity matrix and α is a large number.

where $\Sigma_{\mathbf{uu}} = E\left(\mathbf{u} \cdot \mathbf{u}'\right)$ is the covariance matrix of the random errors.

6 Econometric estimation

I estimated parameters of the model by the Maximum Likelihood Method (MLE). The probability density function of the observed variables $\mathbf{z}_t$, $t = 1,...,n$ can be parameterized by the vector $\boldsymbol{\theta} = \left(\delta,\ a,\ b,\ c,\ \kappa,\ \sigma_\varepsilon,\ \sigma_\eta\right)$ and therefore comes from a "family" $\left\{f\left(\mathbf{Z}_n \mid \boldsymbol{\theta}\right), \boldsymbol{\theta} \in \boldsymbol{\Theta}\right\}$, where $\mathbf{Z}_n = \left(\mathbf{z}_1', ..., \mathbf{z}_n'\right)'$ and the set $\boldsymbol{\Theta} = \left\{x \in R^6 : x_i \in (0,1),\ i = 1,2,\ and\ x_j > 0,\ j = 3,...,6\right\}$ is called as a parameter space. The goal is to find $\boldsymbol{\theta}_0$ in the parameter space $\boldsymbol{\Theta}$ that maximizes the likelihood function L defined as $L(\boldsymbol{\theta}) \equiv f\left(\mathbf{Z}_n \mid \boldsymbol{\theta}\right)$.

How to use results from the Kalman filter in computation of the likelihood function is described for example in Harvey [3], chapter 3.4. The basic idea is that probability density function can be expressed by means of conditional probability densities, which can be easily proved by induction:

$$L(\boldsymbol{\theta}) = \prod_{t=1}^{n} f\left(\mathbf{z}_t \mid \mathbf{Z}_{t-1}, \boldsymbol{\theta}\right). \tag{14}$$

Because random errors are assumed to be normally distributed, the distribution of $\mathbf{z}_t$ conditioned on $\mathbf{Z}_{t-1}$ is commonly known to be also normal and is therefore fully described by its mean vector and covariance matrix. Because of the relation $\mathbf{z}_t = \mathbf{D}\mathbf{x}_t$, these characteristics depend on the conditional mean $\mathbf{x}_{t|t-1}$ and covariance $\mathbf{P}_{t|t-1}$ of the state vector $\mathbf{x}_t$, which was already computed by the Kalman filter recursion. The characteristics of conditional distribution of $\mathbf{z}_t$ are therefore given by the following equations:

$$\mathbf{z}_{t|t-1} = \mathbf{D}\mathbf{x}_{t|t-1}, \tag{15}$$

$$\mathbf{F}_{t|t-1} \equiv E\left[\left(\mathbf{z}_t - \mathbf{z}_{t|t-1}\right)\left(\mathbf{z}_t - \mathbf{z}_{t|t-1}\right)'\right] = \mathbf{D}\mathbf{P}_{t|t-1}\mathbf{D}'. \tag{16}$$

The likelihood function can be therefore calculated as follows:

$$L(\boldsymbol{\theta}) = \prod_{t=1}^{n}\left\{\frac{1}{(2\pi)^{\frac{k}{2}}\left|\mathbf{F}_{t|t-1}\right|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}\left(\mathbf{z}_t - \mathbf{z}_{t|t-1}\right)'\mathbf{F}_{t|t-1}^{-1}\left(\mathbf{z}_t - \mathbf{z}_{t|t-1}\right)\right]\right\}, \tag{17}$$

where k is the number of observable variables.

I maximised the logarithmic form of the likelihood function (17) in Matlab with standard numerical algorithms that are included in fminsearch. Maximization of this likelihood function was, however, complicated by the fact that the function is nearly flat in some directions, which brings us to the concept of identifiability of the parameters.

7 Identifiability

The basic prerequisite for making inference about the value of the vector $\boldsymbol{\theta}$ is that different values of $\boldsymbol{\theta}$ implies different processes according to which data are generated, which Rothenberg [5] formalizes as:

$$\left(f\left(\mathbf{Z}_n \mid \boldsymbol{\theta}\right) = f\left(\mathbf{Z}_n \mid \boldsymbol{\theta}_0\right)\ \ with\ \ probability\ \ 1\right) \Rightarrow \boldsymbol{\theta} = \boldsymbol{\theta}_0. \tag{18}$$

If (18) holds for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, then $\boldsymbol{\theta}_0$ is called to be globally identifiable. If it holds for a nearby neighbourhood of $\boldsymbol{\theta}_0$, then $\boldsymbol{\theta}_0$ is called to be locally identifiable. The model is (locally) identifiable, if every $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$

is (locally) identifiable. If $\boldsymbol{\theta}_0$ is not identifiable, then there is a structure $\boldsymbol{\theta}$ different from $\boldsymbol{\theta}_0$, which cannot be distinguished from $\boldsymbol{\theta}_0$ by observation of the data. Such structures are called observationally equivalent.

From these definitions, one can see that the problem of identifiability arises because of similarity of the probability density functions. It is, however, impossible to deal with the question of identifiability right from these definitions. The userful tool for analysing local identifiability is the following theorem, the proof of which can be found in Rothenberg [5]:

Theorem 1. *Let* $\boldsymbol{\theta}_0$ *be a regular point of the information matrix* $\mathbf{R}_n(\boldsymbol{\theta})$. *Then* $\boldsymbol{\theta}_0$ *is locally identifiable if and only if* $\mathbf{R}_n(\boldsymbol{\theta}_0)$ *is non-singular.*

A point is called regular if it belongs to an open neighborhood where the rank of the matrix does not change. Without this assumption the condition is only sufficient for local identification. The singularity of the information matrix means that the likelihood function is flat at $\boldsymbol{\theta}_0$, which can happen either because some parameters do not affect likelihood at all or different parameters have the same effect on the likelihood function. Iskrev [4] formalized both these situations as follows:

1. changing θ_i does not change the likelihood, so the i-th diagonal element of the information matrix $\Delta_i \equiv E\left(\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_i}\right)^2$ is equal to zero,
2. effect of changing θ_i is offset by changing other parameters at $\boldsymbol{\theta}$, so the multiple correlation coefficient $\rho_i \equiv corr\left(\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_i}; \frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{-i}}\right)$ is equal to one, where $\boldsymbol{\theta}_{-i} \equiv (\theta_1,...,\theta_{i-1},\theta_{i+1},...,\theta_p)$, p being the number of parameters.

Iskrev argues that parameters are unidentified if economic features they represent are either unimportant on their own ($\Delta_i = 0$) or they are redundant given the other features in the model ($\rho_i = 1$). The case of $\rho_i = 1$ then consider as a sign of "overparameterization" of the model, which is particularly relevant for large DSGE models that are often criticized for being too rich in features. From this point of view, theory of identification has important economic modelling aspect and most common suggestion for reducing identifiability problem is to reduce the model. Identification problem can also arise because of data deficiency (not enough or inadequate observable variables, not enough observations), or because of statistical methodology, which is, however, not our case as the method of maximum likelihood uses all the available information about the data generating process, which is described by the probability density function.

Iskrev points out that in most macroeconomic models the information matrix is albeit regular however nearly singular, which is a case that he called weak local identifiability and the reasons for this to happen can be formally written as $\Delta_i \approx 0$ or $\rho_i \approx 1$. By this reasoning he adopted a particular measure of identification

$$s_i^2 \equiv \Delta_i \cdot \left(1 - \rho_i^2\right) \tag{19}$$

and showed that $\left(1/s_i^2\right)$ is equal to the i-th diagonal element of $\left(\mathbf{R}_n(\boldsymbol{\theta})\right)^{-1}$ provided that $\mathbf{R}_n(\boldsymbol{\theta})$ is regular.

How to compute information matrix derives for example Harvey [3] in chapter 3.4. The element in the i-th row and j-th column is computed according to:

$$\left[\mathbf{R}_n(\boldsymbol{\theta})\right]_{ij} = \frac{1}{2}\sum_{t=1}^{T} tr\left(\mathbf{F}_{t|t-1}^{-1}\frac{\partial \mathbf{F}_{t|t-1}}{\partial \theta_i}\mathbf{F}_{t|t-1}^{-1}\frac{\partial \mathbf{F}_{t|t-1}}{\partial \theta_j}\right) + E\left(\sum_{t=1}^{T}\frac{\partial\left(\mathbf{z}_t - \mathbf{z}_{t|t-1}\right)'}{\partial \theta_i}\mathbf{F}_{t|t-1}^{-1}\frac{\partial\left(\mathbf{z}_t - \mathbf{z}_{t|t-1}\right)}{\partial \theta_j}\right). \tag{20}$$

where $\partial \mathbf{F}_{t|t-1}/\partial\theta_i$ and $\partial\left(\mathbf{z}_t - \mathbf{z}_{t|t-1}\right)/\partial\theta_j$ were computed numerically.

The mean value at the right side of this equation is unfortunately impossible to compute except very special cases and so the usual approach is to drop the expectation operator in the second term.

8 Results

The information matrix was nearly singular at all randomly chosen points in the parameter space and the third diagonal element of $\left(\mathbf{R}_n(\boldsymbol{\theta})\right)^{-1}$ corresponding to the parameter b in the model was incomparably higher than other diagonal elements, which indicates very weak identification of this parameter. At first glance, this result seems rather strange as it says that the sensitivity of unemployment to the interest rate is not important in the process of generating data. This can be explained by the fact that the interest rate is not a part of the data set because in the presented model it is seen as a control variable of the central bank and not as an observable indicator of a state variable. Lower sensitivity of unemployment to interest rate is compensated by a more aggressive strategy for setting interest rate by the central bank, which produces practically the same data for unemployment and inflation because of which it is nearly impossible to identify the parameter b in the model. So in this case the problem with identification of parameter b is caused by insufficient number of observable variables.

When treating parameter b as a known constant there were still two relatively large diagonal elements in $\left(\mathbf{R}_n(\boldsymbol{\theta})\right)^{-1}$ at all randomly chosen points in the parameter space and these elements corresponded to the parameters δ and κ in the loss function of the central bank. The problem with weak identification was solved only after treating any of these parameters as a known constant, which suggests that these two parameters play very similar role in the model. This result is rather surprising as economic interpretation of these parameters is quite different. This can be intuitively explained by the fact that in the presented model the inflation is controlled only through unemployment and therefore there is a very close relationship between these two variables, which causes parameters δ and κ to play similar role. Modeling inflation in a more complicated way would thus solve the problem with weak identification of the parameters from another part of the model (from the los function in our case), which can be regarded as analogy to so called identification paradox known from simultaneous equations models. My contribution of this analysis is that weak identification of the parameters (δ and κ in our case) is not necessarily a consequence of "overparameterization" of the model (i.e. that the model is too large) as is frequently argued nowadays in the context of large DSGE models, but a result of omitting important relationships in the model (i.e. that the model is too small). So, in my opinion, the frequently cited suggestion to reduce the model not only needn't reduce the problem with identifiability, but can make things even worse.

When fixing $b = 0.2$ and $\delta = 1$, the other five parameters were estimated by MLE as follows: $a = 0.73$, $c = 1.33$, $\kappa = 0.007$, $\sigma_\varepsilon = 0.57$, $\sigma_\eta = 0.34$. I also performed simulations with these parameter values and generated 1000 data sets each having 100 observations. Then I estimated all seven parameters by MLE 1000-times using standard numerical algorithms based on derivatives when maximizing likelihood function with initial values of the parameters equal to the true ones. The most surprising was that the standard error of the most weakly identified parameter b computed from these simulations was only 0.17 while for example computed standard error for a much better identified parameter a was 0.6. This result does not mean that the true maximum was ordinarily located near the true value of this parameter. In fact, the true maximum was not found at all. The explanation for the obtained result is that the parameter b is very weakly identified and so the likelihood is nearly flat in that direction, which is the reason why numerical algorithm practically did not proceed in this direction. Not to mention the fact that there had to be problems with numerical precision as numerical algorithms based on derivatives use ordinarily inverse of the hessian of the likelihood (or some approximation of it), which corresponds to the (nearly singular) information matrix. These results show that it is not only time-consuming, but also practically impossible to assess the reliability of the estimated parameters using simulation techniques.

Acknowledgements

Financial support of IGA F4/1/2012 is gratefully acknowledged by the author.

References

[1] Ball, L.: Efficient rules for monetary policy. *International Finance*, Wiley Blackwell, vol. 2(1), April, 1999, 63-83.

[2] Chow, G: *Analysis and control of dynamic economic systems.* John Wiley & Sons, New York, 1975.

[3] Harvey, A.C.: *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, 1989.

[4] Iskrev, N.: Evaluating the strength of identification in DSGE models. An a priori approach. W*orking paper w201032*, Banco de Portugal, Economics and Research Department, 2010.

[5] Rothenberg, T.: Identification in parametric models. *Econometrica*, 39, 1971, 577-591.

Stock market speculative bubbles: the case of Visegrad countries

Oleg Deev[1], Veronika Kajurová[2], Daniel Stavárek[3]

Abstract: Conventional theory of speculative bubbles describes stock bubbles as stock prices that exceed their fundamental value because current owners believe that the stocks can be resold at an even higher price in the future. We employ a special methodological technique examine the presence of the phenomenon of stock market bubbles in the Visegrad group countries (Czech Republic, Hungary, Poland, and Slovakia) and selected developed European stock markets. The methodology is based on the examining of residuals of VAR fundamentals with exclusion of ARCH effects. The presence of bubbles is studied by regime switching tests and Hurst persistence tests. Although we examine the bubbles presence over various time periods we found almost no evidence of speculative bubbles across the markets.

Keywords: stock bubble, regime switching test, Hurst persistence test

JEL Classification: G14, G15
AMS Classification: 91G70

1 Introduction

The phenomenon of asset bubbles has been known and studied for centuries, but there is still no common framework on how to detect or predict the formation of a bubble. Since 1980s bubbles are investigated with application of time series econometric analysis. However, the use of such mathematical apparatus raises the question of whether econometric tests can truly detect a bubble or just discover an error in the market evaluation of assets. The majority of empirical research (such as Bohl [4] or Nasseh and Strauss [15]) examines the existence of stock market bubbles using traditional unit root tests of price-dividend ratios of US highly capitalized companies, for which long-term data are available. Other studies were also conducted on markets, for which the history of dividend payments exists. Unfortunately, there are only few papers investigating the occurrence of asset bubbles in emerging markets, especially considering the fact that in the last twenty years those economies were the subjects of large financial inflows and the data on dividends are of limited use. Emerging market studies, primarily focused on China or countries of MENA region (such as Jahan-Parvar and Waters [10], Lehkonen [13] or Ahmed et al [3]), reveal inconsistent results. We can find several references concerning the stock bubbles in Visegrad group countries within some of published papers (such as Kizys and Pierdzioch [11]. The study dealt with the collapse of stock markets in the Czech Republic, Poland and Hungary during the financial crisis and if it was due to international linkages of deteriorating fundamentals or international spillovers of speculative bubbles; Hanousek and Novotný [8] performed an extensive analysis of price jump for emerging stock market indexes from the CEE Visegrad region), but not overall research focused on the stock market bubbles in this region. Clearly new methodological approaches and more research in the area are needed. Furthermore, the individual analysis of the possibility of stock market bubbles in smaller financial markets in Europe, including the Czech Republic and other central European countries, is of particular interest.

The main aim of this paper is to examine the presence of stock market bubbles in the central European countries, namely in the Czech Republic, Hungary, Poland, and Slovakia. In order to prove the statistical significance of proposed methodology, the results are compared to the outcomes of research obtained from selected European developed markets, such as Germany, Austria, France and the United Kingdom. The presence of the 2007-2009 market turmoil brought by the global financial crisis is addressed by dividing the long-time horizon data into three periods: before, during and after the financial crisis.

Identifying stock bubbles is a challenging task not only in terms of time, but also in terms of distinguishing the fundamental and non-fundamental determinants. Since the fundamental value is not directly observable, it must be estimated. On the other hand, it is difficult to confirm the existence of a bubble with a particular

[1] Masaryk University, Faculty of Economics and Administration, Department of Finance, Lipová 41a, 602 00 Brno, oleg@econ.muni.cz.
[2] Masaryk University, Faculty of Economics and Administration, Department of Finance, Lipová 41a, 602 00 Brno, vkajurova@seznam.cz.
[3] Silesian University in Opava, School of Business Administration in Karviná, Department of Finance, Univerzitní nám. 1934/9, 733 40 Karviná, stavarek@opf.slu.cz.

certainty, since the determination of the fundamental value is not a trivial task [12]. The majority of valuation models include dividend payments in their calculations, which values are not available in our inquiry. Therefore, following the methodology of Ahmad et al [1], [2], we assume, that changes in dividends are reflected in the market prices, and abstract from dividends, basing our analysis only on the stock index returns.

2 Data

For the purposes of our study high-frequency daily data are preferred, taking into consideration the market environment of advanced information technology and rapid general information sharing. Daily data capture speedy information as both short run and long run dynamic linkages play role in the bubble formation. We employ the data from the main stock indices representing the chosen markets (basic characteristics of stock exchanges is given in Table 1), weighted indices of the profitability (long-term interest rates) of 10-year government bonds for each country (as calculated in Bloomberg), and the MSCI world index, summarizing the developments of the global stock markets. Returns of the variables as their first log differences are used.

Tick symbol	Country	Number of listed stocks	Market capitalization, mln. USD	Market capitalization as % of GDP	Market turnover, mln. USD	Market turnover as % of GDP	Market liquidity %
PX	Czech Rep (CZ)	16	43 055.6	22.4	14 082.5	7.3	29.4
BUX	Hungary (HU)	48	27 708.4	21.5	26 466.1	20.6	94.5
WIG	Poland (PL)	569	190 234.9	40.5	77 463.9	16.5	47.6
KSM	Slovakia (SK)	90	4 149.6	4.8	173.7	0.2	3.9
ATX	Austria (AU)	86	67 682.8	17.9	48 117.4	12.7	79.4
DAX	Germany (DE)	571	1 429 706.7	43.6	1 405 037.1	42.8	103.0
CAC	France (FR)	901	1 926 488.3	75.3	1 467 073.7	57.3	75.3
UKX	UK	2056	3 107 037.9	137.4	3 006 680.0	132.9	101.9

Sources: The World Bank (World Development Indicators)

Table 1 Stock market characteristics of selected stock exchanges at the end of 2010

The study period is similar for all countries. The total sample period is divided into three sub-periods according to clearly observed trends in the prices' movements. The sub-periods are the pre-crisis period (May 2004 – July 2007), the crisis period (August 2007 – March 2009) and the post-crisis period (April 2009 – March 2012).

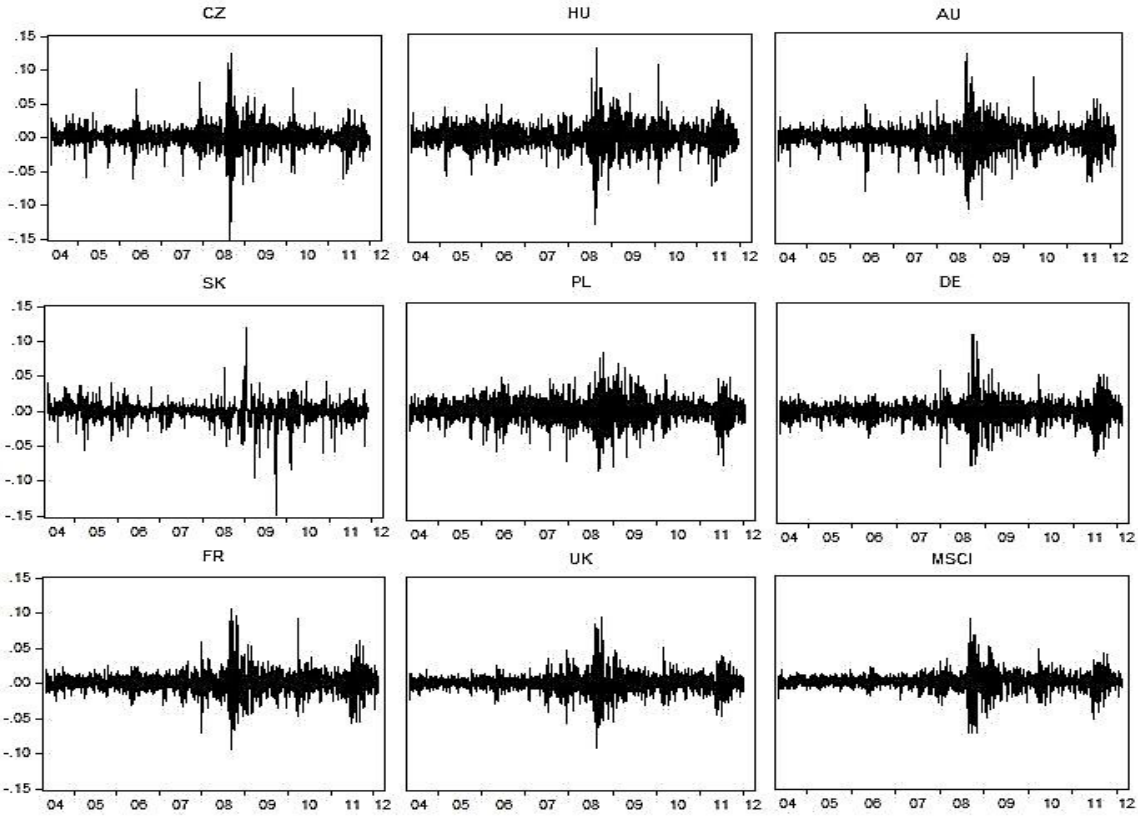


Figure 1 Daily returns (in %) of stock market indices during the whole period

The beginning of the study period is set as an accession date of the Visegrad group countries to the European Union. We chose the crisis period to not start with the Lehman Brothers bankruptcy and major panic on the markets; instead we would like to capture prior anticipations on the markets, when the 2007 banking crisis changed the comfort expectations with a fear of it becoming a sovereign debt crisis. Fluctuations of index returns are illustrated in Figure 1.

3 Methodology

To estimate the fundamental value of index returns we apply the methodology based on the VAR modeling. The VAR model with index returns, government bonds' interest rate and world index returns is employed for each country:

$$r_t = A_1 r_{t-1} + \cdots + A_p r_{t-p} + B msci_t + C b_t + \varepsilon_t \tag{1}$$

where r_t are index returns, $msci_t$ are returns of the MSCI world index, b_t are 10-year government bond interest rates and ε_t is an error term.

To exclude the phenomenon of changing volatility in the time series we should remove autoregressive conditional heteroskedasticity (ARCH) effects from VAR residual series. According to Engle [6] the nonlinear variance dependence measure of ARCH is:

$$\varepsilon_t = \delta_t \mu_t \tag{2}$$

$$\delta_t^2 = \alpha_0 + \sum_{i=1}^{n} \alpha_1 \varepsilon_{1-i}^2 \tag{3}$$

with μ is independent and identically distributed (i.i.d.) variable and α_1 is a coefficient for chosen lags.

Two tests for the identification of stock market speculative bubbles are used. First, we employ the Hurst persistence test to find the existence of long-term linear dependence (memory) in the stock market volatility. Second, we perform the rescale range test. The method based on the Hurst persistence approach is also called rescaled range (R/S) analysis because the significance test breaks the sample into sub-subsamples and then estimates a Chow test on the null that the sub-periods possess identical slopes [1]. This test was developed by Hurst [9] and was firstly implemented in economic analysis by Mandelbrot [14]. Using the R/S analysis, the Hurst exponent H is estimated from the VAR residual series:

$$(R/S)_n = \frac{1}{S(n)}\left[\max_{1\le t\le n} \sum_{t=1}^{n} \left(\varepsilon(t) - \bar{\varepsilon}(n)\right) - \min_{1\le t\le n} \sum_{t=1}^{n} \left(\varepsilon(t) - \bar{\varepsilon}(n)\right)\right] \tag{4}$$

where S_n is the standard deviation estimation and $\bar{\varepsilon}(n)$ is the sample mean of the return time series:

$$\bar{\varepsilon}(\mathrm{n}) = (1/\mathrm{n}) \sum_{\mathrm{n}} \varepsilon(\mathrm{n}) \tag{5}$$

R/S is then described as:

$$(R/S)_n = \left(\frac{n}{2}\right)^H \tag{6}$$

Hurst exponent allows us to reveal the behavior of stock market efficiency over time [16]. If 0 < H < 0.5, it denotes an anti-persistent behavior, which means that positive trends in one period tend to become negative and vice versa. If 0.5 < H < 1, a persistent behavior is indicated in stock market behavior, that is, positive trends in one period tend to continue being positive and vice versa. If H is close to 0.5, it indicates a random walk in data, meaning that market returns are independent. Estimated Hurst exponents are then used to compute F-values for the Chow test to examine its statistical significance.

Second test to detect bubbles in stock market time series is the regime-switching test introduced by Hamilton [7]. The approach of Engle and Hamilton [5] is utilized to test the null hypothesis of no bubbles:

$$\varepsilon_t = trend_t + z_t \tag{7}$$

where z_t is the white noise and $trend_t = \mu_1 + \mu_2 s_t$ (8)

with $s = 1$ being a positive trend and $s = 0$ being a negative trend. Moreover, we let:

$$Prob[s_t = 1\ s_{t-1} = 1] = p, Prob[s_t = 0\ s_{t-1} = 1] = 1 - p \tag{9}$$

$$Prob[s_t = 0\ s_{t-1} = 0] = q, Prob[s_t = 1\ s_{t-1} = 0] = 1 - q \tag{10}$$

The null hypothesis of no trend is given by $p = 1 - q$ and the Wald test statistic calculated as:

$$\frac{p - (1-q)}{var(p) + var(1-q) + covar(p, 1-q)} \tag{11}$$

The Wald test statistic evaluates how close the unrestricted estimates come to satisfying the restrictions under the null hypothesis.

The results of both tests allows us perceiving asset bubbles in the chosen Eastern European countries with a certain degree of confidence, since tests unveils different characteristics of the same time series.

4 Empirical findings

Both tests' results indicate the same situation of no bubbles in stock markets of studied countries (with one exception of Slovakia from the results of Hurst persistence test). There is no significant difference in persistence of stock returns in the highly developed European countries and the Visegrad countries, except for Slovakia. For the majority of cases, Hurst exponent values are not significantly different from its average of 0.5 (see Table 2). Stock index prices follow random walk and do not show any speculation developments.

However, Hurst exponent values for Slovakia highlight irregular market dynamics, which probably disclose the overall inefficiency of the market rather than the existence of price bubbles. The statistical significance of Hurst persistence tests is verified by the Chow test, F-values of which are above its critical values, hence the null hypothesis of no persistence in the time series is rejected.

Estimated Hurst exponents of residuals				
	Full sample	**Pre-crisis period**	**Crisis period**	**Post-crisis period**
PX	0.526192	0.571909	0.508225	0.504513
BUX	0.517590	0.568348	0.418661	0.533226
SKSM	0.653698	0.731891	0.612539	0.531983
WIG20	0.498979	0.496172	0.406587	0.464584
ATX	0.543151	0.557359	0.463424	0.534663
DAX	0.484378	0.466146	0.452908	0.590430
CAC	0.550297	0.486435	0.448344	0.549227
UKX	0.452292	0.515526	0.482410	0.511554
F-values for Chow test				
	Full sample	**Pre-crisis period**	**Crisis period**	**Post-crisis period**
PX	59.093300	60.659000	60.193500	60.400200
BUX	59.603600	56.668500	65.820900	58.693000
SKSM	52.023100	48.131800	54.246500	58.766000
WIG20	60.722600	60.903800	66.618200	62.857500
ATX	58.100000	57.293400	62.944600	58.608400
DAX	61.615200	62.757900	63.608500	55.434000
CAC	57.686500	61.499000	63.898900	57.762200
UKX	63.623400	59.738100	61.763500	59.977100
Critical value F = 4.61				

Source: Authors' calculations based on data from Bloomberg

Table 2 Hurst exponents and related Chow test results

Based on the Hurst exponent values in the sub-periods, the Visegrad stock markets appear to be more volatile than the developed markets in the later periods (except for Poland), supposedly indicating the presence of growing financial inflows. From the results of the Hurst persistence test, the global financial crisis might be seen as a stabilizing mechanism updating the upturning and downgrading market forces (for example, through changing the trading trends in France and Germany). Less than 0.5 values of Hurst exponents in the crisis period signify the decline of asset prices in all observed markets.

Table 3 reports results of regime switching tests. The null hypothesis of no trend in all investigated stock market returns is rejected. Estimated critical value for rejecting the null hypothesis is in all cases lower than the values of the Wald test statistics.

	Full sample	Pre-crisis period	Crisis period	Post-crisis period
PX	819.360	568.246	136.253	304.379
BUX	1123.30	760.251	138.439	441.899
SKSM	791.087	510.217	113.721	265.227
WIG20	1191.91	692.992	236.185	385.211
ATX	707.994	394.937	147.353	290.520
DAX	444.307	480.623	43.4170	136.398
CAC	1552.93	961.337	351.834	780.026
UKX	447.418	324.963	119.101	128.735
Critical value $\chi^2(1) = 3.84$				

Source: Authors' calculations based on data from Bloomberg

Table 3 Wald test results

5 Conclusions

We found no evidence of stock market bubbles neither in the countries of the Visegrad group, nor in the developed European countries. However, taking into consideration the limitations of the proposed methodology, we could not declare with the full certainty that asset bubbles are not present in those markets. If tests have not proved the existence of bubbles, they at least have identified the substantial volatility. Further search of relevant methodology is needed, while tests should be performed not only on market indices, but also on chosen stocks and industry indices.

Acknowledgements

Support of the Czech Science Foundation within the project GAČR 403/11/2073 „Procyclicality of financial markets, asset price bubbles and macroprudential regulation" is gratefully acknowledged.

References

[1] Ahmed, E., Rosser, J. B., Jr., and Uppal, J. Y. Emerging markets and stock market bubbles: Nonlinear speculation? 2010, 1-35.
[2] Ahmed, E., Rosser, J. B., Jr., and Uppal, J. Y. Financial development and bubbles: The case of the Karachi stock exhange of Pakistan, 2010, 1-27.
[3] Ahmed, E., Honggang, L., and Rosser, J. B., Jr. Nonlinear bubbles in Chinese stock markets in the 1990s. *Eastern economic journal* **40** (2006), 1-18.
[4] Bohl, M. T. Periodically collapsing bubbles in the US Stock Market? *International Review of Economics and Finance* **12** (2003), 385-397.
[5] Engle, C., and Hamilton, J. D. Long swings in the Dollar: Are they in the data and do markets know it? *American Economic Review* **80** (1990), 689-713.
[6] Engle, R. F. Autoregressive conditional heteroskedasticity with estimation of the variance of United Kingdom inflation. *Econometrica* **55** (1982), 251-276.
[7] Hamilton, J. D. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **57** (1989), 357-384.
[8] Hanousek, J. and Novotný, J. Price jumps in Visegrad-country stock markets: *An empirical analysis. Emerging Markets Review* (2012), 184-201.
[9] Hurst, H. E. Long term storage capacity of reserviors. *Transformations of the American Society of Civil Engineers* **116** (1951), 770-799.
[10] Jahan-Panvar, M. R., and Waters, G. A. Equity price bubbles in the Middle Eastern and North African financial markets. *Emerging Markets Review* **1** (2010), 39-48.
[11] Kizys, R. and Pierdzioch, Ch. The financial crisis and the stock markets of the CEE countries. *Czech Journal of Economics and Finance* **61** (2011), 153-172.
[12] Kubicová, I., and Komárek, L. The classification and identification of asset price bubbles. *Czech Journal of Economics and Finance* **61** (2011), 34-48.
[13] Lehkonen, H. Bubbles in China. *International Review of Financial Analysis* **19** (2010), 113-117.
[14] Mandelbrot, B. B. Statistical methodology for nonperiodic cycles: From covariance to R/S analysis. *Annals of Economic and Social Measurement* **1** (1972), 259-290.
[15] Nasseh, A., and Strauss, J. Stock prices and the dividend discount model: did their relation break down in the 1990s? *The Quarterly Review of Economics and Finance* **44** (2004), 191-2007.
[16] Rejichi, I. Z., and Aloui, C. Hurst exponent behavior and assessment of the MENA stock markets efficiency. *Research in International Business and Finance* **26** (2012), 353-370.

Efficiency and resource allocation within a hierarchical organization

Martin Dlouhý[1]

Abstract. The objective of production unit is an efficient production by minimizing inputs and maximizing outputs. If the production unit is a part of a larger hierarchical organization the sum of inputs or outputs can be fixed. As example we will use academic departments. The departments are parts of the faculty and the total budget of the faculty is fixed. We will evaluate technical efficiency of the university departments of the Faculty of Informatics and Statistics, University of Economics in Prague. The performance of each department is measured the direct teaching (weighted number of lectures and seminars), indirect teaching (weighted number of exams, theses), and research activity (weighted number of publication points). The objective of the faculty is a fair distribution of resources according to the achieved performance. The zero sum gain DEA model is used for distribution of a fixed budget.

Keywords: Data envelopment analysis, technical efficiency, zero sum gains, university departments.
JEL Classification: C6, I2
AMS Classification: 90C08

1. Introduction

The objective of production unit is an efficient production. This can be achieved by minimizing inputs and maximizing outputs. However, if the production unit is a part of a larger hierarchical organization the sum of inputs or the sum of outputs can be fixed. As an example, we will use efficiency evaluation of academic departments. The departments are parts of the faculty and the total budget of the faculty is fixed. Another interesting example is the Olympic Games country ranking based on the numbers of medals the countries won [7]. Clearly, the number of medals is fixed. A better result of one country can be achieved only by the worse results of the other countries.

The technical efficiency of academic departments is measured by the data envelopment analysis (DEA) that is described in section two. In section three we will evaluate technical efficiency of academic departments first, and then, we will use the special type of the DEA model for setting the budgets of departments.

2. Data Envelopment Analysis

Data envelopment analysis (DEA) evaluates technical efficiency of production unit with the help of mathematical programming. In 1957, Farrell [4] in his paper on the measurement of technical efficiency of production, promoted the ideas to specify the production frontier as the most pessimistic piecewise linear envelopment of the data and to construct efficiency measures based on radial uniform contractions or expansions from inefficient observations to the frontier. The DEA model for multiple inputs and outputs was formulated and solved by Charnes, Cooper, and Rhodes in 1978 [2]. Since 1978 a great variety of DEA models with various extensions and modifications has been developed. These extensions can be found, for example, in textbooks such as Charnes, Cooper, Lewin, and Seiford [1], or Cooper, Seiford, and Tone [3], which present also many examples of applications.

Each production unit allocates its resources into a number of inputs to produce various outputs. DEA uses quantities of inputs consumed and outputs produced to calculate the relative technical efficiency of a production unit. The relative technical efficiency of the unit is defined as the ratio of its total weighted output to its total weighted input or, vice versa, as the ratio of its total weighted input to its total weighted output.

DEA allows each production unit to choose its own weights of inputs and outputs in order to maximize its efficiency score. A technically efficient production unit is able to find such weights that it lies on the production frontier. The production frontier represents the maximum amounts of output that can be produced by given amounts of input (in the output maximization model) or, alternatively, the minimum amounts of inputs required to produce the given amount of output (in the input minimization model).

[1] University of Economics in Prague, Faculty of Informatics and Statistics, Department of Econometrics, e-mail: dlouhy@vse.c.

For each production unit, the DEA model calculates the efficiency score; determines the relative weights of inputs and outputs; and identifies peers for each production unit that is not technically efficient. The peers of a technically inefficient production unit are technically efficient production units with similar combinations of inputs and outputs. The peers serve as benchmarks, which show potential improvements that the technically inefficient production unit can attain. Because the peers are real production units, one can expect that the efficiency improvements should be attainable by the inefficient units.

Now we will continue with a mathematical formulation of the basic DEA model. The mathematical formulation of the input-oriented CCR (Charnes-Cooper-Rhodes) model [2], which assumes the constant returns to scale, is:

$$\begin{aligned}
&\text{maximize} && \frac{\sum_{i=1}^{r} u_i y_{iq}}{\sum_{j=1}^{m} v_j x_{jq}}, \\
&\text{subject to} && \frac{\sum_{i=1}^{r} u_i y_{ik}}{\sum_{j=1}^{m} v_j x_{jk}} \le 1, \quad k = 1, 2, \ldots, n, \qquad (1)\\
& && u_i \ge \varepsilon,\ i = 1, 2, \ldots, r, \\
& && v_j \ge \varepsilon,\ j = 1, 2, \ldots, m,
\end{aligned}$$

where x_{jk} is the amount of input j used by production unit k, y_{ik} is the amount of ith output produced by production unit k, weights u_i and v_j are variables in the DEA model, ε is infinitesimal constant. The value of objective function is the efficiency score of unit q. This program (1) is an input-oriented version of the CCR model. To solve the program (1) by linear programming, the program has to be transformed to the linear form

$$\begin{aligned}
&\text{maximize} && \sum_{i=1}^{r} u_i y_{iq}, \\
&\text{subject to} && \sum_{i=1}^{r} u_i y_{ik} \le \sum_{j=1}^{m} v_j x_{jk}, \quad k = 1, 2, \ldots, n, \qquad (2)\\
& && \sum_{j=1}^{m} v_j x_{jq} = 1, \\
& && u_i \ge \varepsilon,\ i = 1, 2, \ldots, r, \\
& && v_j \ge \varepsilon,\ j = 1, 2, \ldots, m.
\end{aligned}$$

The DEA model (2) has to be formulated and solved for each production unit q=1, 2, ..., n. A specialized software for DEA or MS Excel applications are available at present, which makes it easy to carry out all necessary calculations.

3. Efficiency Evaluation of Academic Departments

We will evaluate technical efficiency of the academic departments of the Faculty of Informatics and Statistics, University of Economics in Prague, in year 2011. The Faculty of Informatics and Statistics is divided into nine academic departments. The performance of each academic department is measured by three outputs: direct

teaching (measured by the weighted number of lectures and seminars), indirect teaching (measured by the weighted number of exams and theses), and research activity (measured by the weighted number of publication points). The only input is the budget of department in 2011. In the definition of inputs and outputs we follow Jablonsky [6], who in his paper also evaluated efficiency of academic departments by a similar method. All input and output data are recalculated in such a way that the faculty as the whole will have input and output values equal to 100. The department data thus show the percentages of three inputs and one output of the faculty that are consumed and produced by each academic department (Table 1). Instead of using the real names of departments, we denote them as departments D1, D2,…, D9.

The obvious objective of each academic department is to use the available resources in order to maximize production of outputs (direct and indirect teaching and research). The total budget of the faculty is fixed and then distributed between departments. This means that the DEA model used for calculation of efficiency should be output-oriented. The input data and results of the output-oriented CCR DEA model are presented in Table 1. Three departments are technically efficient (D3, D4, D7), which means that we are able to find weights ensuring that these three departments use resources in the most efficient way. In the reality, the uniform weights are set by the faculty for each year. However, the weights as well as definitions of performance criteria can be changed, so the technical efficiency can be a better indicator of performance than the results obtained by certain weights that are set for one year.

Department	**Budget**	**Direct Teaching**	**Indirect Teaching**	**Research Activity**	**Efficiency Score**
D1	4.28	4.42	3.85	6.44	0.95
D2	16.02	13.02	12.76	23.12	0.87
D3	2.48	1.88	1.07	6.10	1.00
D4	24.23	24.90	37.58	16.70	1.00
D5	8.72	8.98	7.30	7.86	0.75
D6	12.10	13.50	8.03	8.65	0.73
D7	0.73	1.51	1.30	0.00	1.00
D8	12.23	11.97	14.50	11.65	0.91
D9	19.22	19.82	13.62	19.47	0.76
Total	100.00	100.00	100.00	100.00	

Table 1 Description of inputs and outputs and efficiency score

As we already stated, the objective of the academic department is to use the money in order to maximize production of outputs (teaching and research). This can be seen as efficient use of resources. However, the objective of the department is also to maximize resources for the next period. When the faculty receives its annual faculty budget from the university, the total faculty budget minus the cost of the faculty management is distributed among departments. The total faculty budget is distributed according to performance criteria that include the levels of direct teaching, indirect teaching and research activity. From this view, the teaching and research activity in period one are the inputs and the budget of the department in period two is the output.

Because the departments are parts of the faculty, the total faculty budget to be distributed among departments is fixed. The money additionally allocated to one department means that budgets of other departments have to be reduced. For this type of resource allocation, the so-called zero sum gains DEA model (ZSG-DEA) was developed [7]. The idea of this model is to re-allocate inputs or outputs in such a way that all production units (in this case academic departments) will become technically efficient. For the output-oriented ZSG-DEA model, the following equation is valid:

$$h_{ZSGi} = h_{DEAi}\left(1 - \frac{\sum_{j \in W} y_j \left[\left(h_{DEAi} / h_{DEAj}\right) h_{ZSGi} - 1\right]}{\sum_{j \notin W} y_j}\right),$$

where h_{ZSGi} is the efficiency score of production unit i in the ZSG-DEA model, h_{DEAi} is the efficiency score of production unit i in the traditional DEA model, W is a set of inefficient production units.

In the ZSG-DEA model with one output (in this case the budget of department), the ZSG-DEA model can be simplified. Suppose that the faculty will allocate a constant budget B among n departments based on given inputs (direct and indirect teaching and research activity). In the initial DEA model, regardless of the input level, the amount B/n is allocated to each department. Then we solve a DEA model with the single constant output and three inputs. The final projection $h_{DEAi}/\Sigma h_{DEAi}$ is ZSG-DEA efficient. This solution can be shown to be invariant under scale and orientation of the DEA model [5]. According to our view, the ZSG-DEA solution is much easier than the iteration procedure used by Jablonsky [6].

The results of the output-oriented traditional DEA model and the results of the simplified output-oriented ZSG-DEA model are presented in Table 2. The *initial efficiency* shows the technical efficiency of each academic department with the same output level (constant budget). In this case, 11.11% of the total faculty budget is allocated to each academic department. In the proposed allocation, the higher budgets (outputs) are allocated to more efficient academic departments and the lower budgets are allocated to less efficient departments in comparison to the original budgets of departments presented in Table 1. Hence the *proposed budget* assures that all academic departments are technically efficient (Table 2). This can be checked by calculating DEA model with the original values of teaching and research activity as inputs and the proposed budget as an output. As one can see, *final efficiency* of all academic departments equals to one in such model (Table 2).

Department	Constant Budget	Direct Teaching	Indirect Teaching	Research Activity	Efficiency Initial	Proposed Budget	Efficiency final
D1	11.11	4.42	3.85	6.44	2.93	4.78	1.00
D2	11.11	13.02	12.76	23.12	8.64	14.10	1.00
D3	11.11	1.88	1.07	6.10	1.00	1.63	1.00
D4	11.11	24.90	37.58	16.70	16.53	26.96	1.00
D5	11.11	8.98	7.30	7.86	5.77	9.41	1.00
D6	11.11	13.50	8.03	8.65	6.44	10.50	1.00
D7	11.11	1.51	1.30	0.00	1.00	1.63	1.00
D8	11.11	11.97	14.50	11.65	7.95	12.96	1.00
D9	11.11	19.82	13.62	19.47	11.06	18.03	1.00
Total	100.00	100.00	100.00	100.00	x	100.00	x

Table 2 Results of the ZSG-DEA Model

4. Conclusion

If the production unit is a part of a larger hierarchical organization, the sum of inputs or the sum of outputs can be fixed. As an example, we showed this on the efficiency evaluation of academic departments. The academic departments are parts of the faculty and the total budget of the faculty is fixed. A special type of the DEA model known as the ZSG-DEA model can be used for the setting budgets of academic departments in such a way that all production units (departments) become technically efficient.

Acknowledgements

The research is supported by the Czech Science Foundation - project no. P403/12/1387 "Models for Efficiency Evaluation in Hierarchical Economic Systems".

References

[1] Charnes, A., Cooper, W. W., Lewin, A. Y, Seiford, L. M. (Eds.): *Data envelopment analysis: theory, methodology and applications*, Boston: Kluwer Academic Publisher, 1994.

[2] Charnes, A., Cooper, W. W., Rhodes, E.: Measuring the inefficiency of decision making units. *European Journal of Operational Research* **2** (1978), 429–444.

[3] Cooper, W. W., Seiford, L.W., Tone, K.: *Data envelopment analysis*. Boston: Kluwer Academic Publisher, 2000.

[4] Farell, M. J.: The measurement of productive efficiency of production. *Journal of the Royal Statistical Society*, Series A, **120 (III)** (1957), 253–281.

[5] Gomes, E. G., Da Silva e Souza, G.: Allocating financial resources for competitive projects using a zero sum gains DEA model. *Engevista* **12** (2010), 4-9.

[6] Jablonsky, J.: Models for efficiency evaluation in education. *8th International Conference on Efficiency and Responsibility in Education*. Czech University of Life Sciences Prague, 2011, pp. 110-119.

[7] Lins, M. P. E., Gomes, E. G., Soares De Mello, J. C. C. B., Soares De Mello, A. J. R.: Olympic ranking based on a zero sum gains DEA model. *European Journal of Operational Research* **148** (2003), 312-322.

About a modification of $E_r/E_s/1/m$ queueing system subject to breakdowns

Michal Dorda[1], Dušan Teichmann[2]

Abstract. The paper deals with modelling of a finite single-server queueing system with a server subject to breakdowns. Customers interarrival times and customers service times follow the Erlang distribution defined by the shape parameter r or s and the scale parameter $r\lambda$ or $s\mu$ respectively. We consider that server failures can occur when the server is either idle or busy (so called operate-independent failures). Further we assume that service of a customer is interrupted by the occurrence of the server failure and the system empties when the server is broken (we call it as the failure-empty discipline). We assume that random variables relevant to server failures and repairs are exponentially distributed. We use two approaches - analytical approach using method of stages and simulation approach using coloured Petri nets. We use both approaches to compare and validation created models. At the end of the paper some reached results are shown.

Keywords: $E_r/E_s/1/m$, queueing, breakdowns, disasters, Petri net.

JEL Classification: C44
AMS Classification: 60K25

1 Introduction

Queuing theory is a tool which enables us to find characteristics of queueing systems. We meet queueing systems for example in informatics, transport and economics. In general, a queueing system represents a system which serves customers coming in the system. For a lot of queuing systems which were solved in the past it is assumed that there are no failures of servers. Such queueing systems are often called reliable queueing systems. The second part of queueing systems are represented by so called unreliable queueing systems or queueing systems subject to server breakdowns.

One of the first queueing systems subject to breakdowns was solved by Avi-Itzhak and Naor [1]. Today we can find a lot of papers solving queuing systems subject to breakdowns. As regards papers written during last 20 years we can mention following papers. Lam et al. [3] modelled a single-server queue with a repairable server under the assumption of the Poisson arrival process and exponentially distributed service times. Tang [7] published the paper devoted to an unreliable single-server queue as well, but with generally distributed service times. Sharma and Sirohi [5] modelled a container unloader as a finite single-server queue with repairable server.

Some queues with several unreliable servers were studied for example by following authors. Martin and Mitrani [4] studied a system with several unreliable servers placed in parallel. Wang and Chang [8] considered a finite multi-server queue with balking, reneging and server breakdowns.

Interesting group of unreliable queueing systems is formed by queues with so-called negative customers or disasters (or catastrophes). Disasters can represent server failures which cause removing either some or all customers finding in the system. We can for example mention the papers written by Boxma et al. [2] or Shin [6].

On the basis of the short review of queueing systems subject to breakdowns we can state that most of the authors studied especially unreliable queueing systems under the assumptions of the exponential or general distribution. Most of the mentioned authors further assumed that a queue of waiting customers has an infinity capacity, if the queue of waiting customers is formed.

In the paper we will pay our attention to a finite single-server queueing system with the server subject to breakdowns, where customers interarrival times and service times will follow the Erlang distribution. Further we will assume that times between failures and times to repair will be exponential random variables. We use the Erlang distribution because it offers us greater variability of usage than the exponential distribution, nevertheless

[1] VŠB – Technical University of Ostrava /Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, Ostrava – Poruba, tel. number: +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, tel. number: +420 597 324 575, e-mail: dusan.teichmann@vsb.cz.

mathematical models of queueing systems in which we assume the Erlang distribution are still relatively easily solvable.

The paper is organized as follows. In Section 2 we will make necessary assumptions. In Section 3 we will present the mathematical model, in Section 4 we will present a simulation model of studied queueing system. Section 5 is devoted to the executed numerical experiments and in Section 6 we make some conclusions.

2 General assumptions and notations

Let us study a single server queueing system with a finite capacity equal to m, where $m>1$, that means there are in total m places for customers in the system – single place in the service and m-1 places intended for waiting of customers. Let us consider that customers are served one by one according to the FCFS service discipline.

Let customers interarrival times follow the Erlang distribution with the shape parameter r and the scale parameter $r\lambda$; therefore the mean interarrival time is then equal to $\frac{r}{r\lambda}=\frac{1}{\lambda}$. Costumer service times are an Erlang random variable with the shape parameter s and the scale parameter $s\mu$; thus the mean service time is equal to $\frac{s}{s\mu}=\frac{1}{\mu}$.

Let us assume that the server is successively failure-free (or available we can say) and under repair. We assume that failures of the server can occur when the server is idle or busy – we say that server failures are operate-independent. Let us assume that times between failures are an exponential random variable with the parameter η; the mean time between failures is then equal to the reciprocal value of the parameter η. Times to repair are an exponential random variable as well, but with the parameter ζ; the mean time to repair is therefore equal to $\frac{1}{\zeta}$. It is clear, that the server steady-state availability A (the fraction of time the server is available) is equal to:

$$A=\frac{\zeta}{\eta+\zeta}$$

and the server steady-state unavailability U (the fraction of time the server is broken) is:

$$U=1-A=\frac{\eta}{\eta+\zeta}.$$

As regards behaviour of customers at the moment of the failure, we will consider the system empties after every failure of the server; the system is empty when the server is down – failures represent disasters in the system.

3 Mathematical model

To model the studied queueing system we applied method of stages. The method exploits the fact that the Erlang distribution with the shape parameter r or s and the scale parameter denoted as $r\lambda$ or $s\mu$ is sum of s or r independent exponential distribution with the same parameter $s\lambda$ or $r\mu$. Therefore the queue can be modelled by Markov chains using.

Let us consider a random variable $K(t)$ being the number of the customers finding in the system, a random variable $I(t)$ being the number of terminated phases of customer arrival, a random variable $J(t)$ being the number of terminated phases of customer service and a random variable $F(t)$ being the number of broken servers at the time t. On the basis of the assumptions established in Section 2 it is clear that $\{K(t), I(t), J(t), F(t)\}$ constitutes a multi-dimensional Markov process with the state space

$$\Omega=\{(k,i,j,f),k=0,i=0,...,r-1,j=0,f=0,1\}\cup\{(k,i,j,f),k=1,...,m,i=0,...,r-1,j=0,...,s-1,f=0\}.$$

The system is found in the state (k,i,j,f) at the time t if $K(t)=k$, $I(t)=i$, $J(t)=j$ and $F(t)=f$, let us denote the corresponding probability $P_{(k,i,j,f)}(t)$.

Let us illustrate the queueing model graphically as a state transition diagram (see figure 1). The vertices represent the particular states of the system and oriented edges indicate the possible transitions with the corresponding rate. Please notice that in figure 1 there are depicted only selected states that are necessary for formation of the equation system.

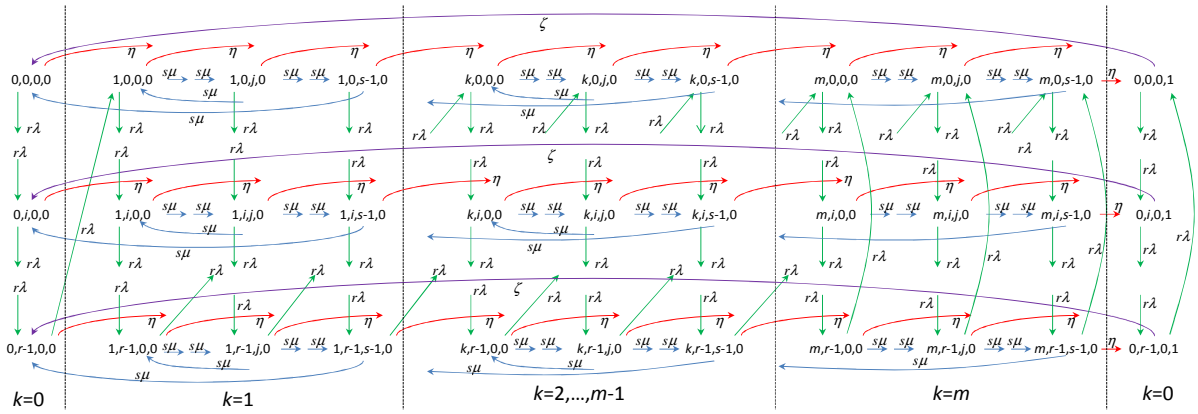


Figure 1 The state transition diagram

On the basis of the state transition diagram we are able to write a finite linear equation system of steady-state balance equations in the form:

$$(r\lambda+\eta)\cdot P_{(0,0,0,0)} = s\mu\cdot P_{(1,0,s-1,0)} + \zeta\cdot P_{(0,0,0,1)}, \tag{1}$$

$$(r\lambda+\eta)\cdot P_{(0,i,0,0)} = r\lambda\cdot P_{(0,i-1,0,0)} + s\mu\cdot P_{(1,i,s-1,0)} + \zeta\cdot P_{(0,i,0,1)} \text{ for } i=1,\ldots,r-1, \tag{2}$$

$$(r\lambda+s\mu+\eta)\cdot P_{(k,0,0,0)} = r\lambda\cdot P_{(k-1,r-1,0,0)} + s\mu\cdot P_{(k+1,0,s-1,0)} \text{ for } k=1,\ldots,m-1, \tag{3}$$

$$(r\lambda+s\mu+\eta)\cdot P_{(k,i,0,0)} = r\lambda\cdot P_{(k,i-1,0,0)} + s\mu\cdot P_{(k+1,i,s-1,0)} \text{ for } k=1,\ldots,m-1, i=1,\ldots,r-1, \tag{4}$$

$$(r\lambda+s\mu+\eta)\cdot P_{(1,0,j,0)} = s\mu\cdot P_{(1,0,j-1,0)} \text{ for } j=1,\ldots,s-1, \tag{5}$$

$$(r\lambda+s\mu+\eta)\cdot P_{(k,i,j,0)} = r\lambda\cdot P_{(k,i-1,j,0)} + s\mu\cdot P_{(k,i,j-1,0)} \text{ for } k=1,\ldots,m, i=1,\ldots,r-1, j=1,\ldots,s-1, \tag{6}$$

$$(r\lambda+s\mu+\eta)\cdot P_{(k,0,j,0)} = r\lambda\cdot P_{(k-1,r-1,j,0)} + s\mu\cdot P_{(k,0,j-1,0)} \text{ for } k=2,\ldots,m-1, j=1,\ldots,s-1, \tag{7}$$

$$(r\lambda+s\mu+\eta)\cdot P_{(m,0,0,0)} = r\lambda\cdot P_{(m-1,r-1,0,0)} + r\lambda\cdot P_{(m,r-1,0,0)}, \tag{8}$$

$$(r\lambda+s\mu+\eta)\cdot P_{(m,0,j,0)} = r\lambda\cdot P_{(m-1,r-1,j,0)} + r\lambda\cdot P_{(m,r-1,j,0)} + s\mu\cdot P_{(m,0,j-1,0)} \text{ for } j=1,\ldots,s-1, \tag{9}$$

$$(r\lambda+s\mu+\eta)\cdot P_{(m,i,0,0)} = r\lambda\cdot P_{(m,i-1,0,0)} \text{ for } i=1,\ldots,r-1, \tag{10}$$

$$(r\lambda+\zeta)\cdot P_{(0,0,0,1)} = r\lambda\cdot P_{(0,r-1,0,1)} + \eta\cdot P_{(0,0,0,0)} + \eta\cdot\sum_{k=1}^{m}\sum_{j=0}^{s-1} P_{(k,0,j,0)}, \tag{11}$$

$$(r\lambda+\zeta)\cdot P_{(0,i,0,1)} = r\lambda\cdot P_{(0,i-1,0,1)} + \eta\cdot P_{(0,i,0,0)} + \eta\cdot\sum_{k=1}^{m}\sum_{j=0}^{s-1} P_{(k,i,j,0)} \text{ for } i=1,\ldots,r-1 \tag{12}$$

including normalization equation in the form:

$$\sum_{i=0}^{r-1}\sum_{f=0}^{1} P_{(0,i,0,f)} + \sum_{k=1}^{m}\sum_{i=0}^{r-1}\sum_{j=0}^{s-1} P_{(k,i,j,0)} = 1. \tag{13}$$

We have the equation system of $m\cdot r\cdot s+2r+1$ linear equations formed by equations (1) up to (13) with $m\cdot r\cdot s+2r$ unknown stationary probabilities. To solve it we can omit for example equation (1). Solving of the system we executed using Matlab. Applied state description in the form of (*k*,*i*,*j*,*f*) is four-dimensional and is very good for formation the equation system but is unsuitable for computations in Matlab. Therefore we established an alternative one-dimensional state description in the following form:

- The states (*k*,*i*,*j*,*f*) for *k*=1,...,*m*, *i*=0,...,*r*-1, *j*=0,...,*s*-1 and *f*=0 can be denoted using a single value $(k-1)\cdot r\cdot s + j\cdot r + i + 1$,
- The states (*k*,*i*,*j*,*f*) for *k*=0, *i*=0,...,*r*-1, *j*=0 and *f*=0,1 can be denoted using a single value $m\cdot r\cdot s + f\cdot r + i + 1$.

Applying the alternative one-dimensional state description we are able to transform the equation system in the form we need for using Matlab (we need a transition matrix). After numerical solving of the equation system rewritten in matrix form we obtain the stationary probabilities we need in order to compute performance meas-

ures of the studied system. Let us consider three performance measures – the mean number of the customers in the service *ES*, the mean number of the customers waiting in the queue *EL* and the mean number of the broken servers *EP*.

For the performance measures we can write following formulas:

$$ES = \sum_{k=1}^{m}\sum_{i=0}^{r-1}\sum_{j=0}^{s-1} P_{(k,i,j,0)},$$

$$EL = \sum_{k=2}^{m}(k-1)\cdot\sum_{i=0}^{r-1}\sum_{j=0}^{s-1} P_{(k,i,j,0)},$$

$$EP = \sum_{i=0}^{r-1} P_{0,i,0,1} = U .$$

4 Simulation model

In order to validate the outcomes, which were reached by solution of the above-mentioned mathematical model, Petri net model of the studied queueing system was created by using CPN Tools – Version 3.0.4. The software CPN Tools is designed for editing, simulating and analyzing coloured Petri nets. The created simulation model in initial marking is shown in figure 2. The model is compound of 11 places and 10 transitions. The Petri net presented in figure 2 models the unreliable $E_r/E_s/1/5$ queueing system fulfilling the conditions mentioned in Section 2. In figure 2 the applied values of individual parameters are r=2, $r\lambda$=18 h^{-1}, s=2, $s\mu$=20 h^{-1}, η=0.1 $h^{-1}$and ζ=0.2 h^{-1}.

The concrete values of the random variables are generated during the simulation through the defined function *fun ET(k, mi) = round(erlang(k, mi/3600.0))*; where *k* is the shape parameter and *mi* is the scale parameter of the Erlang probability distribution expressed in [h^{-1}]. We apply a second as the unit of time.

The created Petri net works with following tokens:

- Timed tokens c represent customers.
- Timed tokens f represent failures of the server.
- Auxiliary tokens p serve for modelling for example free queue places or free servers.

To obtain desired simulation outcomes three monitoring functions were defined:

- The monitoring function named *ES* which is bound with the place *Busy servers* enables estimation of the mean number of the customers in the service.
- The monitoring function named *EL* which is bound with the place *Queue* serves for estimation of the mean number of the waiting customers.
- The monitoring function named *EP* which is bound with the place *Repairing* was create in order to estimate the mean number of the broken servers.

In order to stop the simulation after reaching defined simulation time we created a breakpoint function which stops each simulation run after reaching the simulation time 31 536 000 s; the value corresponds to 365 days. And finally, to execute 30 simulation runs for each experiment we defined an auxiliary text *CPN'Replications.nreplications 30*. Evaluating the text defined number of simulation runs is performed and simulation outcomes gained.

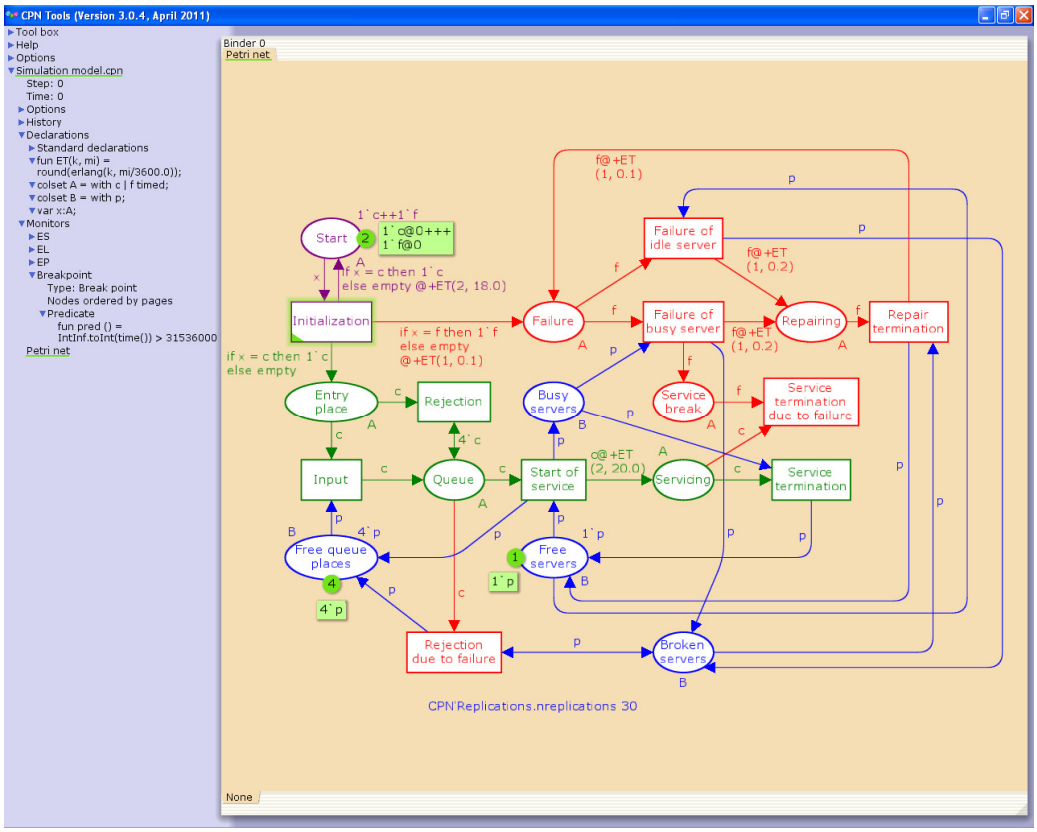


Figure 2 Simulation model created in CPN Tools – Version 3.0.4

5 Outcomes of executed experiments

We executed several experiments with both models to compare analytical and simulation outcomes and to obtain some graphical dependencies. Applied values of model parameters are summarized in table 1.

Parameter	m [-]	r [-]	$r\lambda$ [h^{-1}]	s [-]	$s\mu$ [h^{-1}]	η^{-1} [h]	ζ [h^{-1}]
Applied value	5	2	18	2	20	10 up to 200 with step 10	0.2

Table 1 Applied values of model parameters

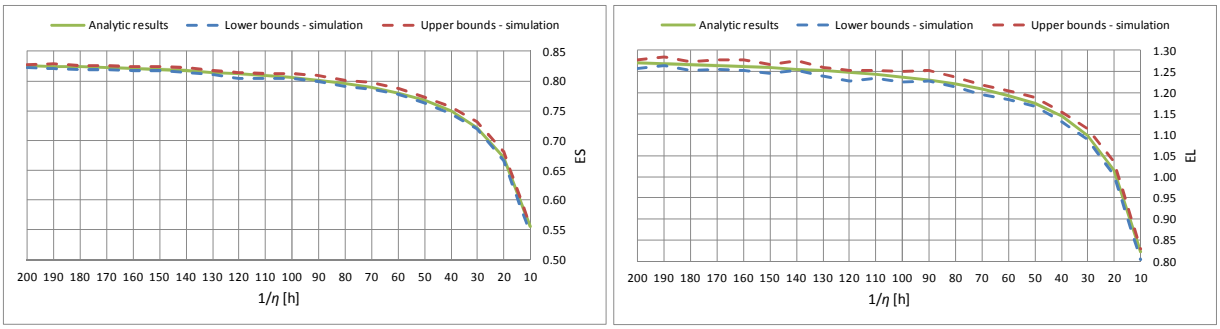


Figure 3 The graphical dependency of *ES* and *EL* on the reciprocal value of η

In table 2 we can see the comparison of analytical and simulation outcomes. On the basis of executed simulations we computed 95% confidence intervals for *ES*, *EL* and *EP* where T_l is the lower bound and T_u is the upper bound of the interval. We can see that for all executed experiments the analytical value of individual performance measures lies in the corresponding confidence interval. We see that both models gives very similar outcomes, we validated the mathematical model.

The graphical dependencies of *ES* and *EL* on η^{-1} are shown in figure 3. As can be seen, both dependencies are decreasing; the fact could be logically expected because increasing η (or decreasing η^{-1}) means more frequent

failures that cause lower fraction of time in which the server is able to serve incoming customers and the system is often empty.

η^{-1}	Analytical results			Simulation results					
	ES	*EL*	*EP*	T_l for *ES*	T_u for *ES*	T_l for *EL*	T_u for *EL*	T_l for *EP*	T_u for *EP*
200	0.82593	1.27017	0.02439	0.82170	0.82775	1.25782	1.27871	0.02220	0.02892
190	0.82483	1.26835	0.02564	0.82103	0.82869	1.26434	1.28446	0.02316	0.03079
180	0.82362	1.26633	0.02703	0.81918	0.82614	1.25314	1.27232	0.02375	0.03171
170	0.82226	1.26408	0.02857	0.81879	0.82553	1.25546	1.27814	0.02601	0.03242
160	0.82075	1.26156	0.03030	0.81826	0.82474	1.25185	1.27700	0.02592	0.03208
150	0.81903	1.25872	0.03226	0.81701	0.82351	1.24636	1.26739	0.02692	0.03403
140	0.81708	1.25548	0.03448	0.81457	0.82254	1.25267	1.27641	0.02935	0.03833
130	0.81484	1.25177	0.03704	0.81089	0.81695	1.23865	1.25899	0.03467	0.04087
120	0.81224	1.24746	0.04000	0.80426	0.81486	1.22653	1.25160	0.03764	0.04836
110	0.80920	1.24240	0.04348	0.80450	0.81252	1.23368	1.25310	0.03924	0.04820
100	0.80557	1.23637	0.04762	0.80425	0.81242	1.22531	1.25105	0.03994	0.04770
90	0.80117	1.22908	0.05263	0.79867	0.80864	1.22795	1.25334	0.04702	0.05719
80	0.79575	1.22008	0.05882	0.79148	0.80062	1.21378	1.23613	0.05415	0.06423
70	0.78887	1.20868	0.06667	0.78569	0.79764	1.19535	1.21796	0.05479	0.06839
60	0.77988	1.19377	0.07692	0.77824	0.78744	1.18335	1.20488	0.06809	0.07870
50	0.76762	1.17345	0.09091	0.76243	0.77290	1.16723	1.18747	0.08532	0.09770
40	0.74992	1.14411	0.11111	0.74482	0.75587	1.13000	1.15255	0.10409	0.11654
30	0.72211	1.09804	0.14286	0.71973	0.73204	1.08981	1.11331	0.13153	0.14577
20	0.67207	1.01525	0.20000	0.66656	0.68169	1.00384	1.03585	0.18984	0.20624
10	0.55541	0.82288	0.33333	0.54409	0.55807	0.80376	0.82949	0.33059	0.34610

Table 2 Outcomes of executed experiments

6 Conclusions

In the paper we presented the mathematical model of $E_r/E_s/1/m$ queueing system subject to disasters which causes loss of all customers finding in the system and rejection of all customers incoming to the system while the server is under repair. The mathematical model was solved using Matlab to get the stationary probabilities of system states. The probabilities we need for computing performance measures. We focused on three performance measures *ES*, *EL* and *EP*. The analytic outcomes were validated using simulation; we created a simple simulation model of the system using coloured Petri net. As regard our future research we would like to extend the model; there are another performance measures we are interested in. For example we would like to compute the probability that customer will be rejected upon his arrival or due to the failure of the server.

References

[1] Avi-Itzhak B., Naor P.: Some queueing problems with the service station subject to breakdown. *Operations Research* **11** (1963); 303-320.

[2] Boxma O. J., Perry D., Stadje W.: Clearing models for M/G/1 queues. *Queueing Systems* **38** (2001); 287-306.

[3] Lam Y., Zhang Y. L., Liu Q.: A geometric process model for M/M/1 queueing system with a repairable service station. *European Journal of Operational Research* **168** (2006); 100-121.

[4] Martin S. P., Mitrani I.: Analysis of job transfer policies in systems with unreliable servers. *Annals of Operations Research* **162** (2008); 127-141.

[5] Sharma K. C., Sirohi A.: Cost analysis of the unloader queueing system with a single unloader subject to breakdown with two types of trailers. *Opsearch* **47**(1) (2010); 93-103.

[6] Shin Y. W.: Multi-server retrial queue with negative customers and disasters. *Queueing Systems* **55** (2007); 223-237.

[7] Tang Y. H.: A single server M/G/1 queueing system subject to breakdowns - some reliability and queueing problems. *Microelectronics Reliability* **37**(2) (1997); 315-321.

[8] Wang K., Chang Y.: Cost analysis of a finite M/M/R queueing system with balking, reneging, and server breakdowns. *Mathematical Methods of Operations Research* **56** (2002); 169-180.

Output analysis and stress testing for mean-variance efficient portfolios

Jitka Dupačová[1]

Abstract. Solutions of portfolio optimization problems are often influenced by model misspecifications or by errors due to approximation, estimation and incomplete information. The obtained results, recommendations for the risk and portfolio manager should be then carefully analyzed. We shall focus on output analysis and stress testing with respect to uncertainty or perturbations of input data for the Markowitz mean-variance model with a general polyhedral convex set of considered portfolios and we shall discuss its robust versions. Possible extensions to general mean-risk efficient portfolios will be delineated.

Keywords: Markowitz model, mean-variance efficient portfolios, stability, stress testing, worst-case analysis.

JEL classification: D81, G11, C61
AMS classification: 91G10, 90C31

1 The Markowitz model

In conclusions of his famous paper [14] on portfolio selection, Markowitz stated that "what is needed is essentially a 'probabilistic' reformulation of security analysis". He developed a model for portfolio optimization in an uncertain environment under various simplifications. It is a static, single period model which assumes a frictionless market. It applies to small rational investors whose investments cannot influence the market prices and who prefer higher yields to lower ones and smaller risks to larger ones. Let us recall the basic formulation: The composition of portfolio of I assets is given by weights of the considered assets, $x_i, i = 1, \dots, I$, $\sum_i x_i = 1$. The unit investment in the i-th asset provides the random return ρ_i over the considered fixed period. The assumed probability distribution of the vector ρ of returns of all assets is characterized by a known vector of expected returns $E\rho = \mu$ and by a fixed covariance matrix $V = [\mathrm{cov}(\rho_i, \rho_j),\, i, j = 1, \dots, I]$ whose main diagonal consists of variances of individual returns. This allows to quantify the "yield from the investment" as the expectation $\mu(x) = \sum_i x_i \mu_i = \mu^\top x$ of its total return and the "risk of the investment" as the variance of its total return, $\sigma^2(x) = \sum_{i,j} \mathrm{cov}(\rho_i, \rho_j) x_i x_j = x^\top V x$. According to the assumptions, the investors aim at maximal possible yields and, at the same time, at minimal possible risks – hence, a typical decision problem with two criteria, "max" $\{\mu(x), -\sigma^2(x)\}$ or "min" $\{-\mu(x), \sigma^2(x)\}$. The mean-variance efficiency introduced by Markowitz is fully in line with general concepts of multiobjective optimization. Accordingly, mean-variance efficient portfolios can be obtained by solving various optimization problems such as

$$\min_{x \in \mathcal{X}} \{-\lambda \mu^\top x + 1/2 x^\top V x\} \tag{1}$$

where the value of parameter $\lambda \geq 0$ reflects investor's risk aversion. Another possibility, favored in the practice, is to minimize the portfolio variance subject to a lower bound for the total expected return, i.e.

$$\min_{x \in \mathcal{X}} x^\top V x \text{ subject to } \mu^\top x \geq k \tag{2}$$

with parameter k, or to maximize the expected return under a constraint on portfolio variance

$$\max_{x \in \mathcal{X}} \mu^\top x \text{ subject to } x^\top V x \leq v. \tag{3}$$

In the classical theory, the set $\mathcal{X} = \{x \in \mathbb{R}^I : \sum_i x_i = 1\}$ without nonnegativity constraints, which means that short sales are permitted. Under this simplification explicit forms of optimal solutions can

[1]Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, Prague 8, 18675, Czech Republic, e-mail: dupacova@karlin.mff.cuni.cz

be obtained, e.g. the optimal solution of (1) $x(\mu, V; \lambda)$ is linear in μ. We shall allow for general convex polyhedral sets $\mathcal{X}$. To trace out the mean-variance frontier, one may solve (1), (2) or (3) for many different values of λ, k, v respectively, or to rely on parametric programming techniques, cf. [16]. Notice that the same set of mean-variance solutions is obtained when using $\sqrt{x^\top V x}$ at the place of $x^\top V x$.

It was the introduction of risk into the investment decisions which was the exceptional feature of this model and a real breakthrough, and the Markowitz model became a standard tool for portfolio optimization. However, there are many questions to be answered: Modeling the random returns to get their expectations, variances and covariances, the choice of the value of λ, etc. From the point of view of optimization procedures, inclusion of nonnegativity or *linear* regulatory constraints does not cause any serious problems. This however does not apply to minimal transaction unit constraints which introduce 0-1 variables. In the interpretation and application of the results one has to be aware of the model assumptions (not necessarily fulfilled in real-life), namely, that it is a one-period model based on the buy-and-hold strategy applied between the initial investment and the horizon of the problem so that decisions based on its repeated use over more that one period can be far from a good, suboptimal dynamic decision, cf. [3]. See also [17] and references therein for a discussion and multiperiod extensions.

2 Output analysis for the Markowitz mean-variance model

The optimal solution $x(\mu, V\,; \lambda)$ and the optimal value $\varphi(\mu, V\,; \lambda)$ of (1) depend on μ, V and on the chosen value of λ, and at the same time, one can hardly assume full knowledge of these input values. The impact of errors in expected returns, variances and covariances on the optimal return φ of the obtained portfolio was investigated, e.g. in [5]: The program (1) was solved repeatedly with perturbed selected input parameters, ceteris paribus, and the cash equivalent loss was computed for each run. The results of this simulation study indicate that the errors in expected values are more influential than those in the second order moments. Inspired by the cited results we shall deal first with sensitivity analysis of the optimal composition of the portfolio and of the optimal value of (1) on the input values of the expected returns μ of the risky assets and we shall suggest to complement results based on parametric programming by *stochastic sensitivity analysis*.

Assume that the covariance matrix V in (1) is a known *positive definite* matrix, the set $\mathcal{X}$ a nonempty convex polyhedron with nondegenerated vertices, $\lambda > 0$ a chosen parameter value, and that the expected return μ is a parameter of the quadratic program (1). The covariance matrix V and the parameter λ will not be indicated in our denotation of the optimal value and of the optimal solution of (1). Under the above assumptions, for each μ, there is a unique optimal solution $x(\mu)$ of (1) and the optimal value function $\varphi(\mu) := \min_{x\in\mathcal{X}}[-\lambda\mu^\top x + \frac{1}{2}x^\top V x]$ is a concave function. This follows from a more general statement which is a direct consequence of the inequality valid for minimum of a sum:

Proposition 1. *Assume that the objective function $f : \mathcal{X} \times \mathbb{R}^q \to \mathbb{R}$ in the parametric program*

$$\min_{x\in\mathcal{X}} f(x, p)$$

is linear in the parameter p, the set $\mathcal{X}$ is a non-empty convex set which does not depend on p, and an optimal solution $x(p)$ exists for all p. Then the optimal $\varphi(p)$ is a concave function on $\mathbb{R}^q$.

The set of feasible solutions $\mathcal{X}$ of the quadratic program (1) can be decomposed into finitely many relatively open facets that are identified by indices of active constraints; interior of $\mathcal{X}$ and vertices of $\mathcal{X}$ are special cases of these facets. The parametric space $\mathbb{R}^n$ of vectors $p := \lambda\mu$ can be decomposed into finitely many disjoint stability sets linked with the facets by the requirement that for all p belonging to a stability set, the optimal solutions $x(p)$ of the quadratic program (1) lie in the same facet. It is possible to prove (see [2]) that $x(p)$ is continuous on the whole space $\mathbb{R}^n$, is linear on each stability set and differentiable on its interior.

If, however, p belongs to the boundary of a stability set, $x(p)$ looses the differentiability property and is only directionally differentiable. The optimal value function $\varphi(p)$ is piecewise linear – quadratic differentiable concave function of p. These results explain the observed cases of a relative stability of the optimal value and of an extremal sensitivity of optimal solutions on small changes of the vector μ of expected returns: Whenever the initial value of $p = \lambda\mu$ belongs to the boundary of a stability set, arbitrarily small changes in μ can cause transition to one of the neighboring stability sets. It means not

only that some other assets are included into portfolio, but different small changes can cause transition to different stability sets. As a result, the composition of the optimal portfolio is regarded unstable. At the same time, the change of the minimal value of (1) is small for small changes of μ.

Illustrative example. Consider the quadratic program minimize $-p_1x_1 - p_2x_2 + 1/2x_1^2 + x_1x_2 + x_2^2$ on the set $\mathcal{X} = \{x_1 \geq 0, x_2 \geq 0 \ : \ x_1 + x_2 \leq 1\}$ which corresponds to the risk-adjusted expected return problem (1) with 3 assets where short sales are not permitted. Decomposition of $\mathcal{X}$ into facets Σ_k, $k = 1, \ldots, 7$ is depicted in Figure 1a and the corresponding stability sets $\sigma(\Sigma_k), k = 1, \ldots, 7$ are drawn on Figure 1b. Consider point $p_1 = p_2 = 1$ on Figure 1b. For this parameter value, the optimal

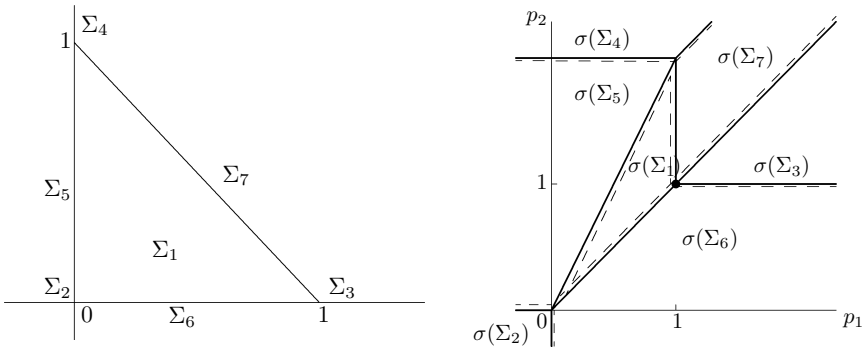


Figure 1: a. Decomposition of the set $\mathcal{X}$ / b. Stability sets

solution is vertex Σ_3, however, a small change of parameter values causes moving the optimal solution into the adjacent facets Σ_6 or Σ_7 or into the interior Σ_1 of $\mathcal{X}$. The corresponding changes of the optimal value and of the first component of the optimal solution are illustrated for fixed $p_2 = 1$ in Figure 2. A similar situation can be observed also in case of changes of the parameter λ (i.e., when tracing the

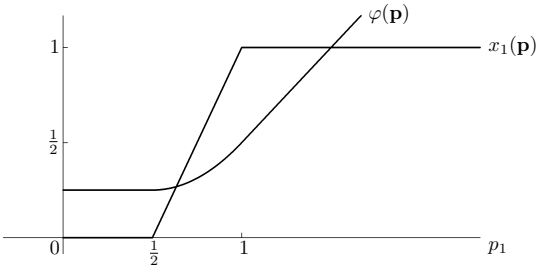


Figure 2: $\varphi(\boldsymbol{p})$ and $x_1(\boldsymbol{p})$ for $p_2 = 1$

mean-variance efficient frontier) but they are more easy to take in as the changes concern only a scalar parameter. Moreover, there is $\lambda_1 > 0$ such that all parameter values $\lambda > \lambda_1$ belong to the same stability set characterized by fixed indices of positive components of efficient portfolio $x(\lambda)$ and that they are linear functions of λ on (λ_1, ∞).

There exist generalizations of the cited results to the case of V positive semidefinite and bounded convex polyhedral set $\mathcal{X}$, hovewer, the fact that from the point of view of quadratic programming there might be multiple optimal solutions indicates clearly the limitations.

Sample based return averages μ_ν are frequently used at the place of the true expectation μ_0. From asymptotic normality of μ_ν, asymptotic normality of the sample optimal value function follows and asymptotic confidence intervals for the optimal value can be constructed.

The optimal, mean-variance efficient solutions $x(\mu)$ are continuous, piece-wise linear, directionally differentiable on certain nonoverlapping convex stability sets in $\mathbb{R}^I$; cf. [2]. Their continuity is sufficient for consistency of the optimal solutions based on a consistent estimate of the true expected return. Except for the simple case of $\mathcal{X} = \{x \in \mathbb{R}^I : \sum_i x_i = 1\}$, $x(\mu)$ are not differentiable. It means that their asymptotic normality holds true only if the true expected return μ_0 lies in the interior of a stability set.

We refer to [8] for details.

3 Stress testing the parameter values

A special type of output analysis appears under the name *stress testing* in the context of quantification of losses or risks which may appear under special, mostly extremal circumstances. Usually, the model is solved or its solution evaluated for an alternative input. We shall indicate now how it is possible to quantify such "stress testing" results.

To stress the parameter values in the Markowitz model we shall apply the contamination technique of robust statistics. In its basic form it requires that the objective function, say $f(x,p)$, is linear with respect to the parameter p and that the set of feasible decisions is fixed, hence, Proposition 1 applies. Changes of parameter p are modeled as $p_t = (1-t)p + t\hat{p}$ where $\hat{p}$ is a selected parameter perturbation to be tested and $0 \leq t \leq 1$ is a scalar parameter. This approach can be applied to (1) with $\mu_t = (1-t)\mu + t\hat{\mu}$, $V_t = (1-t)V + t\hat{V}$ and to (2) or (3) when a *known expected return* μ or a *fixed* covariance matrix V are assumed. To stress separately correlations one can adapt a suggestion of [13]: The covariance matrix can be written as $V = DCD$ with the diagonal matrix D of "volatilities" (standard deviations of the marginal distributions) and the correlation matrix C. Changes in the covariances may be then modeled by "stressing" the correlation matrix C by a positive semidefinite *stress correlation matrix* $\hat{C}$

$$C(\gamma) = (1-\gamma)C + \gamma\hat{C} \tag{4}$$

with parameter $\gamma \in [0,1]$. This type of perturbation of the initial quadratic program allows us to apply the related stability results of [2] to the perturbed problem (2)

$$\min_{x\in\mathcal{X}} x^\top DC(\gamma)Dx,\ \gamma \in [0,1]: \tag{5}$$

where the constraint $\mu^\top x \geq k$ has been incorporaded into the definition of $\mathcal{X}$. Proposition 1 can be specified as

Proposition 2. *Under the above assumptions, the optimal value $\varphi(\gamma)$ of (5) is concave and continuous in $\gamma \in [0,1]$ and the optimal solution $x(\gamma)$ is a continuous vector in the range of γ where $C(\gamma)$ is positive definite.*

Application of [12], Theorem 17, provides the form of the directional derivative

$$\varphi'(0^+) = x^\top(0)D\hat{C}Dx(0) - \varphi(0).$$

Contamination bounds

$$(1-\gamma)x^\top(0)DCDx(0) + \gamma x^\top(1)D\hat{C}Dx(1) \leq \min_{x\in\mathcal{X}} x^\top DC(\gamma)Dx$$

$$\leq (1-\gamma)x^\top(0)DCDx(0) + \gamma x^\top(0)D\hat{C}Dx(0)$$

quantify the effect of the considered change in the input data on the optimal value $\varphi(\gamma)$ of portfolio; cf. [10]. In a similar way, one can quantify the influence of stressing parameters μ, C or μ, V in (1) or parameter μ in (3).

4 Worst-case analysis for the Markowitz mean-variance model

Incomplete knowledge of input data, i.e. of expected returns μ and covariance matrix V may be also approached via the worst-case analysis or robust optimization, cf. [11], [15], [18]. The idea is to hedge against the worst possible input belonging to a prespecified uncertainty or ambiguity set $\mathcal{U}$. We shall denote $\mathcal{M}$, $\mathcal{V}$ considered uncertainty sets for parameters μ and V and will assume that $\mathcal{U} = \mathcal{M} \times \mathcal{V}$. For (1) this means to solve

$$\min_{x\in\mathcal{X}} \max_{(\mu,V)\in\mathcal{U}} \{-\lambda\mu^\top x + 1/2x^\top Vx\}. \tag{6}$$

The worst-case reformulations of (2) and (3) are

$$\min_{x\in\mathcal{X}} \max_{V\in\mathcal{V}} x^\top Vx \text{ subject to } \min_{\mu\in\mathcal{M}} \mu^\top x \geq k, \tag{7}$$

$$\min_{x\in\mathcal{X}} \min_{\mu\in\mathcal{M}} \mu^\top x \text{ subject to } \max_{V\in\mathcal{V}} x^\top V x \le v, \tag{8}$$

respectively. Consider for example $\mathcal{U}$ described by box constraints $0 \le \underline{\mu}_i \le \mu_i \le \overline{\mu}_i$, $i = 1, \dots, I$, $\underline{V} \le V \le \overline{V}$ componentwise and such that V is positive definite. With $\mathcal{X} = \{x \in \mathbb{R}^I : x_i \ge 0\, \forall i, \sum_i x_i = 1\}$ the inner maximum in (6) is attained for $\mu_i = \underline{\mu}_i\, \forall i$ and $V = \overline{V}$. The robust mean-variance portfolio is the optimal solution of

$$\min_{x\in\mathcal{X}} \{-\lambda \underline{\mu}^\top x + 1/2 x^\top \overline{V} x\}.$$

We refer to [11] for a survey of various choices of uncertainty sets for the Markowitz model.

5 General mean-risk portfolios

Let us proceed now to static mean-risk models of the Markowitz type with random returns ρ whose probability distribution P does not depend on the selected portfolio composition. (Recall the assumption of a small investor in the Markowitz model.) The yield from the portfolio x is again the expectation $E_P \rho^\top x$, the risk is understood now as a function R which assigns a real number to the uncertain outcome $\rho^\top x$ for the decision x. The value of function R should not depend on the realization of the uncertain return ρ but it depends on the decision and on the probability distribution P; accordingly we shall denote it $R(x, P)$. It should posses some natural properties such as monotonicity, translation equivariance, positive homogeneity and subadditivity for to be called coherent; see [1]. The well-known risk measure Value at Risk (VaR), which is not coherent in general, and the coherent Conditional Value at Risk (CVaR) are special cases of R.

For a known probability distribution P of returns the problems corresponding to (1), (2), (3) are

$$\min_{x\in\mathcal{X}} \{-\lambda E_P \rho^\top x + R(x, P)\}, \tag{9}$$

$$\min_{x\in\mathcal{X}} R(x, P) \text{ subject to } E_P \rho^\top x \ge k, \tag{10}$$

$$\max_{x\in\mathcal{X}} E_P \rho^\top x \text{ subject to } R(x, P) \le v. \tag{11}$$

The form (9) with a probability independent set of feasible decisions is more convenient for applications of quantitative stability analysis techniques, whereas risk management regulations ask frequently for satisfaction of risk constraints with a fixed limit v displayed in (11). Moreover, (11) is favored in practice: solving it for various values of v one obtains directly the corresponding points $[\mu^\top x(v), v]$ on the mean-risk efficient frontier. Numerical tractability of the mean-risk problems depends on the choice of the risk measure and on the assumed probability distribution P. Programs (9)–(11) are convex for convex risk measures $R(\bullet, P)$, such as CVaR; see [6], [10]. As the probability distribution P is fully known only exceptionally, there are two main tractable ways for analysis of the output regarding changes or perturbation of P – quantitative stability analysis with respect to changes of P by stress testing via contamination, see [6], [7], [9], [10], or the worst-case analysis with respect to all probability distributions belonging to an uncertainty set $\mathcal{P}$ which will be briefly discussed below.

The "robust" counterpart of (9) is a straightforward transcription of (6):

$$\min_{x\in\mathcal{X}} \max_{P\in\mathcal{P}} \{-\lambda E_P \rho^\top x + R(x, P)\} \tag{12}$$

whereas for (11) we have

$$\max_{x\in\mathcal{X}} \min_{P\in\mathcal{P}} E_P \rho^\top x \text{ subject to } R(x, P) \le v \text{ for all } P \in \mathcal{P} \text{ or } \max_{P\in\mathcal{P}} R(x, P) \le v. \tag{13}$$

For the Markowitz model, one in fact considers an uncertainty set of probability distributions characterized by fixed expectations and covariance matrices; the Markowitz model does not distinguish among probability distributions belonging to this set. Accordingly, let us specify the class $\mathcal{P}$ as the class of probability distributions identified by fixed moments μ, V known from the Markowitz model. Thanks to the assumed linearity of random returns explicit formulas for the worst-case CVaR and VaR can be derived, cf. [4], and according to Theorem 2.2 of [18] the portfolio composition $x \in \mathcal{X}$ satisfies the worst-case constraint on VaR iff it satisfies the worst-case CVaR constraint.

In general, for convex, compact classes $\mathcal{P}$ defined by moment conditions and for fixed x, the maxima in (12), (13) are attained at extremal points of $\mathcal{P}$. Then under modest assumptions it is possible to pass in (12) and in (13) to discrete distributions $P \in \mathcal{P}$. This convenient property carries over also to $R(x, P)$ that are *convex* in P.

Whereas expected utility functions or $\mathrm{CVaR}(x, P)$ are linear in P, various popular risk measures are not even convex in P: the variance is concave in P, the mean absolute deviation is neither convex nor concave in P. This means that extensions of the minimax approach to risk functionals nonlinear in P are carried through only under special circumstances.

Acknowledgement. The research was partly supported by the project of the Czech Science Foundation P402/12/G097 "DYME/Dynamic Models in Finance".

References

[1] Artzner, P., Delbaen, J., Eber, J., and Heath, D.: Coherent measures of risk, *Matematical Finance* **9** (1999), 202–228.

[2] Bank, B., Guddat, J., Klatte, D., Kummer, B., and Tammer, K.: *Non-Linear Parametric Optimization.* Akademie-Verlag, Berlin, 1982.

[3] Cariño, D. R., Myers, D. H., and Ziemba, W. T.: Concepts, technical issues and uses of the Russell - Yasuda Kasai financial planning model, *Oper. Res.* **46** (1998), 450–462.

[4] Čerbáková, J.: Worst-case VaR and CVaR. In: *Operations Research Proceedings 2005* (H. D. Haasis, H. Kopfer, and J. Schönberger, eds.), Springer, Berlin, 2006, 817–822.

[5] Chopra, W. K. and Ziemba, W. T.: The effect of errors in means, variances and covariances on optimal portfolio choice, *J. of Portfolio Mgt.* **19** (1993), 6–11.

[6] Dupačová, J.: Stress testing via contamination. In: *Coping with Uncertainty* (K. Marti et al. eds.), **LNEMS 581**, Springer, 2006, 29–46.

[7] Dupačová, J.: *Portfolio Optimization and Risk Management via Stochastic Programming.* Osaka University Press, 2009.

[8] Dupačová, J., Hurt, J., and Štěpán, J.: *Stochastic Modeling in Economics and Finance*, Kluwer, Dordrecht, 2002.

[9] Dupačová, J. and Kopa, M.: Robustness in stochastic programs with risk constraints, *Ann. Oper. Res.* published online 2011, DOI 10.1007/s10479-010-0824-9.

[10] Dupačová, J. and Polívka, J.: Stress testing for VaR and CVaR, *Quantitative Finance* **7** (2006), 411–421.

[11] Fabozzi, F. J., Huang, D., and Zhou, G.: Robust portfolios: Contributions from operations research and finance, *Ann. Oper. Res.* **176** (2010), 191–220.

[12] Gol'shtein, E. G.: *Vypukloje Programmirovanije. Elementy Teoriji.* Nauka, Moscow, 1970. [*Theory of Convex Programming*, Translations of Mathematical Monographs, **36**, American Mathematical Society, Providence RI, 1972].

[13] Kupiec, P.: Stress testing in a Value at Risk framework. In: *Risk Management: Value at Risk and Beyond* (M. A. H. Dempster ed.), Cambridge Univ. Press, 2002, 76–99.

[14] Markowitz, H. M.: Portfolio Selection, *J. of Finance* **6** (1952), 77–91.

[15] Pflug, G. and Wozabal, D.: Ambiguity in portfolio selection, *Quantitative Finance* **7** (2006),435–442.

[16] Ruszczyński, A. and Vanderbei, R. J.: Frontiers of stochastically nondominated portfolios, *Econometrica* **71** (2003), 1287–1297.

[17] Steinbach, M.C.: Markowitz revisited: Mean-variance models in financial portfolio analysis, *SIAM Review* **43** (2001), 31–85.

[18] Zymler, S., Kuhn, D., and Rustem, B.: Distributionally robust joint chance constraints with second-order moment information. *Math. Program. A*, published online, DOI 10.1007/s10107-011-0494-7.

Efficient Score Test for Change Detection in Vector Autoregressive Models

Marek Dvořák[1]

Abstract. The statistics and econometrics articles both contain a vast amount of testing procedures related to structural changes with unknown break dates, most of them designed for the univariate models. Work on detecting change-points in multivariate models is receiving an increasing attention nowadays. Many of the test statistics are based on likelihood ratio, see for instance paper by [5]. We will discuss the efficient score-based test for change detection in multivariate AR models. Its asymptotic properties and results from the simulation study will be presented. The idea of such methodology came from [3].

Keywords: Change point analysis, Invariance principle, VAR

JEL classification: C12, C32
AMS classification: primary 62G20; secondary 60F17, 62M10

1 Introduction

1.1 The model and its assumptions

Let us consider the following $\mathbb{VAR}(p)$ model of the form

$$\mathbf{y}_t = \mathbf{c} + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \ldots + \mathbf{\Phi}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad t \in \mathbb{Z}, \tag{1}$$

where $\mathbf{c} \in \mathbb{R}^n$ is a vector of constants, $\mathbf{\Phi}_j \in \mathbb{R}^{n\times n}$, $j = 1, \ldots, p$ are non random matrices of autoregressive coefficients and $\{\boldsymbol{\varepsilon}_t\}_t$ is an n-dimensional noise process specified further. Let us assume that we have T consecutive observations $\mathbf{y}_1, \ldots, \mathbf{y}_T$ drawn from the model (1). Our objective is to test the null hypothesis H_0 against H_1, where

$$\begin{aligned} H_0 :& \quad \mathbf{y}_t = \mathbf{c} + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \ldots + \mathbf{\Phi}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t,\ t = 1, \ldots, T, \\ H_1 :\ \exists k \in \{1, \ldots, T-1\} :& \quad \mathbf{y}_t = \mathbf{c} + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \ldots + \mathbf{\Phi}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t,\ t = 1, \ldots, k, \\ & \quad = \widetilde{\mathbf{c}} + \widetilde{\mathbf{\Phi}}_1 \mathbf{y}_{t-1} + \ldots + \widetilde{\mathbf{\Phi}}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t,\ t = k+1, \ldots, T \end{aligned} \tag{2}$$

and either $\mathbf{c} \neq \widetilde{\mathbf{c}}$ or $\mathbf{\Phi}_j \neq \widetilde{\mathbf{\Phi}}_j$ for at least one $j = 1, \ldots, p$. We will not assume any changes in the variance structure of the autoregressive model.

Let us denote $\mathbf{I}_n \in \mathbb{R}^{n\times n}$ an n dimensional identity matrix. We define an indicator $\mathrm{I}_{[s=t]}$ with the values 1 for $s = t$ and 0 otherwise. Throughout the paper, vec operator is considered as a matrix function that stacks column vectors of a matrix below one another.

Let us formulate the following assumptions:

Assumptions:

(A.1) roots of the polynomial $\det\{\mathbf{I}_n - \mathbf{\Phi}_1 z - \ldots - \mathbf{\Phi}_p z^p\}$ lie outside the complex unit circle,

(A.2) white noise process $\boldsymbol{\varepsilon}_t$ is i.i.d. Gaussian, $\mathsf{E}[\boldsymbol{\varepsilon}_s \boldsymbol{\varepsilon}_t^\top] = \mathrm{I}_{[s=t]} \mathbf{\Omega}$ and $\mathbf{\Omega}$ is a positive-definite matrix,

[1]Charles University, Faculty of Mathematics and Physics/Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Prague, Czech Republic, e-mail: dvorakm@karlin.mff.cuni.cz

(A.3) vector $\mathbf{v}_1 = \mathsf{vec}(\mathbf{y}_0, \mathbf{y}_{-1}, \ldots, \mathbf{y}_{1-p})$ of initial observations satisfies

$$\mathbf{v}_1 = \boldsymbol{\mu_{np}} + \sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{u}_{-k},$$

where $\boldsymbol{\mu_{np}} := \mathsf{vec}(\boldsymbol{\mu}, \boldsymbol{\mu}, \ldots, \boldsymbol{\mu}) \in \mathbb{R}^{np}$, $\mathbf{u}_t := \mathsf{vec}(\boldsymbol{\varepsilon}_t, \mathbf{0}, \ldots, \mathbf{0}) \in \mathbb{R}^{np}$ and

$$\mathbf{A} = \begin{pmatrix} \boldsymbol{\Phi}_1 & \cdots & \boldsymbol{\Phi}_{p-1} & \boldsymbol{\Phi}_p \\ \mathbf{I}_{p-1} \otimes \mathbf{I}_n & & & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{np \times np}.$$

Assumption (A.1) guarantees stationarity of the $\mathbb{VAR}(p)$ model. Assumption (A.2) specifies the properties of the error term $\boldsymbol{\varepsilon}_t$. The last Assumption (A.3) ensures that the initial observations of the model come from the same stationary representation as the observations $\mathbf{y}_1, \ldots, \mathbf{y}_k$.

1.2 The test statistic

We start with the construction of the efficient vector test statistic that will be derived from the likelihood function under the assumptions above. Let us denote $\boldsymbol{\phi} := \mathsf{vec}\,(\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p) \in \mathbb{R}^{n^2 p}$, $\boldsymbol{\theta} := (\boldsymbol{\mu}^\top, \boldsymbol{\phi}^\top)^\top \in \mathbb{R}^{n(np+1)}$, $\mathbf{Y}_k := \mathsf{vec}(\mathbf{y}_1, \ldots, \mathbf{y}_k) \in \mathbb{R}^{nk}$, $\boldsymbol{\mu_k} := \mathsf{vec}(\boldsymbol{\mu}, \ldots, \boldsymbol{\mu}) \in \mathbb{R}^{nk}$ and

$$\mathbf{X}_k := \begin{pmatrix} \mathbf{y}_0 - \boldsymbol{\mu} & \cdots & \mathbf{y}_{k-1} - \boldsymbol{\mu} \\ \vdots & \ddots & \vdots \\ \mathbf{y}_{1-p} - \boldsymbol{\mu} & \cdots & \mathbf{y}_{k-p} - \boldsymbol{\mu} \end{pmatrix} \in \mathbb{R}^{np \times k}. \tag{3}$$

The conditional log-likelihood function based on k observations $\mathbf{y}_1, \ldots, \mathbf{y}_k$ with given $\mathbf{y}_{1-p}$, $\mathbf{y}_{2-p}, \ldots, \mathbf{y}_0$ is of the form

$$\ell_k(\boldsymbol{\mu}, \boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p, \boldsymbol{\Omega}) = -\frac{nk}{2}\log(2\pi) - \frac{k}{2}\log|\boldsymbol{\Omega}| - \frac{1}{2}\sum_{t=1}^{k} \boldsymbol{\varepsilon}_t^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \boldsymbol{\mu} - \sum_{j=1}^{p} \boldsymbol{\Phi}_j (\mathbf{y}_{t-j} - \boldsymbol{\mu})$. Partial derivatives of ℓ_k with respect to the unknown parameters $\boldsymbol{\mu}$ and $\boldsymbol{\phi}$ are of the form

$$\frac{\partial}{\partial \boldsymbol{\mu}} \ell_k(\boldsymbol{\theta}) = \Big(\mathbf{I}_n - \sum_{j=1}^{p} \boldsymbol{\Phi}_j\Big)^\top \boldsymbol{\Omega}^{-1} \sum_{t=1}^{k} \boldsymbol{\varepsilon}_t,$$

$$\frac{\partial}{\partial \boldsymbol{\phi}} \ell_k(\boldsymbol{\theta}) = \left(\mathbf{X}_k \otimes \boldsymbol{\Omega}^{-1}\right)\left(\mathbf{Y}_k - \boldsymbol{\mu_k}\right) - \left(\mathbf{X}_k \mathbf{X}_k^\top \otimes \boldsymbol{\Omega}^{-1}\right)\boldsymbol{\phi}$$

and "$\otimes$" denotes the Kronecker product.

The Fisher information matrix $\mathcal{I}_{\mathcal{F}}$ about the parameter vector $\boldsymbol{\theta}$ can be obtained under general regularity conditions as

$$\mathcal{I}_{\mathcal{F}}(\boldsymbol{\theta}) = -\,\mathsf{E}\left[\frac{\partial^2 \ell_k}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^\top}\right],$$

where

$$\frac{\partial^2 \ell_k}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^\top} = \begin{pmatrix} -k\Big(\mathbf{I}_n - \sum_{j=1}^{p} \boldsymbol{\Phi}_j\Big)^\top \boldsymbol{\Omega}^{-1}\Big(\mathbf{I}_n - \sum_{j=1}^{p} \boldsymbol{\Phi}_j\Big) & \frac{\partial^2 \ell_k}{\partial \boldsymbol{\mu}\, \partial \boldsymbol{\phi}^\top} \\ \frac{\partial^2 \ell_k}{\partial \boldsymbol{\phi}\, \partial \boldsymbol{\mu}^\top} & -\mathbf{X}_k \mathbf{X}_k^\top \otimes \boldsymbol{\Omega}^{-1} \end{pmatrix}$$

and the off-diagonal elements can be found in [6], p. 91. The asymptotic information matrix $\mathcal{I}(\boldsymbol{\theta}) = \lim_{k\to\infty} \frac{1}{k}\mathcal{I}_{\mathcal{F}}(\boldsymbol{\theta})$ for $\boldsymbol{\theta}$ exists and is given by

$$\mathcal{I}(\boldsymbol{\theta}) = \begin{pmatrix} \Big(\mathbf{I}_n - \sum_{j=1}^{p} \boldsymbol{\Phi}_j\Big)^\top \boldsymbol{\Omega}^{-1}\Big(\mathbf{I}_n - \sum_{j=1}^{p} \boldsymbol{\Phi}_j\Big) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_y(0) \otimes \boldsymbol{\Omega}^{-1} \end{pmatrix} =: \begin{pmatrix} \mathcal{I}_{1,1}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_{2,2}(\boldsymbol{\theta}) \end{pmatrix},$$

see [6], p. 91–92; and $\boldsymbol{\Gamma}_y(0)$ is the variance matrix of $\mathbf{y}_1, \ldots, \mathbf{y}_p$.

Let $\widehat{\boldsymbol{\theta}}_T := \big(\widehat{\boldsymbol{\mu}}_T^\top, \widehat{\boldsymbol{\phi}}_T^\top\big)^\top$ be the maximum likelihood (=ML) estimators of the unknown parameters based on the full sample $\mathbf{y}_1, \ldots, \mathbf{y}_T$. The efficient score test statistic is an $r := n(np+1)$-vector of the following form

$$\widehat{\mathbf{B}}(\tau) := \frac{1}{\sqrt{T}} \cdot \mathcal{I}^{-\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_T) \cdot \begin{pmatrix} \frac{\partial}{\partial \boldsymbol{\mu}} \ell_{\lfloor T\tau \rfloor}(\widehat{\boldsymbol{\theta}}_T) \\ \frac{\partial}{\partial \boldsymbol{\phi}} \ell_{\lfloor T\tau \rfloor}(\widehat{\boldsymbol{\theta}}_T) \end{pmatrix},$$

where $\lfloor x \rfloor$ is integer part of x and $k = \lfloor T\tau \rfloor$, $0 \le \tau \le 1$.

Throughout the article, we will omit the subscript $_T$ in the ML estimators for notation simplicity. Let us denote $\mathbf{K} = \sum_{j=1}^p \boldsymbol{\Phi}_j$ and its corresponding ML estimate as $\widehat{\mathbf{K}} = \sum_{j=1}^p \widehat{\boldsymbol{\Phi}}_j$.

2 Main Result

The aim of the article will be the proof of the following theorem:

Theorem 1. *Let us suppose that the sequence $\{\mathbf{y}_t\}$ satisfies $\mathbb{VAR}(p)$ model of the form (1) and Assumptions (A.1) – (A.3) be fulfilled. Then, under H_0, there exists a r-dimensional sequence of Brownian bridges $\mathbf{B}(\tau)$ with independent components $B_j(\tau)$, $0 \le \tau \le 1$, $j = 1, \ldots, r$, such that*

$$\max_{j=1,\ldots,r} \sup_{0\le\tau\le 1} \big|\widehat{B}_j(\tau) - B_j(\tau)\big| = o_{\mathsf{P}}(1). \tag{4}$$

Proof. The proof can be done separately with respect to each parameter and the basic steps are similar to those in [3]. Due to the consistency of the ML estimators, it follows that $\big\|\mathcal{I}(\widehat{\boldsymbol{\theta}}) - \mathcal{I}(\boldsymbol{\theta})\big\| = o_{\mathsf{P}}(1)$.

(i) First we consider the case of detecting change in $\boldsymbol{\mu}$. The component of the efficient score vector that we use is

$$\frac{1}{\sqrt{T}}\left[\frac{\partial}{\partial \boldsymbol{\mu}}\ell_k(\boldsymbol{\theta})\right]_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = \frac{1}{\sqrt{T}}\big(\mathbf{I}_n - \widehat{\mathbf{K}}\big)^\top \widehat{\boldsymbol{\Omega}}^{-1} \sum_{t=1}^k \Big(\mathbf{y}_t - \widehat{\boldsymbol{\mu}} - \sum_{j=1}^p \widehat{\boldsymbol{\Phi}}_j(\mathbf{y}_{t-j} - \widehat{\boldsymbol{\mu}})\Big) + \tag{5}$$

$$+ \frac{1}{\sqrt{T}}\big(\mathbf{I}_n - \mathbf{K}\big)^\top \boldsymbol{\Omega}^{-1} \sum_{t=1}^k \Big(\mathbf{y}_t - \overline{\mathbf{y}} - \sum_{j=1}^p \boldsymbol{\Phi}_j(\mathbf{y}_{t-j} - \overline{\mathbf{y}})\Big) - \tag{6}$$

$$- \frac{1}{\sqrt{T}}\big(\mathbf{I}_n - \mathbf{K}\big)^\top \boldsymbol{\Omega}^{-1} \sum_{t=1}^k \Big(\mathbf{y}_t - \overline{\mathbf{y}} - \sum_{j=1}^p \boldsymbol{\Phi}_j(\mathbf{y}_{t-j} - \overline{\mathbf{y}})\Big), \tag{7}$$

where $\overline{\mathbf{y}} := T^{-1}\sum_{t=1}^T \mathbf{y}_t$. We will show that the second addend (6) converges to the Brownian bridge process and also that the difference between (5) and (7) is negligible. This will complete the first part of the proof. Let M be a positive constant which can be different from term to term. For the second addend (6) we obtain that

$$\sup_{0\le\tau\le 1}\left\|\frac{1}{\sqrt{T}}\big(\mathbf{I}_n - \mathbf{K}\big)^\top \boldsymbol{\Omega}^{-1} \sum_{t=1}^{\lfloor T\tau\rfloor} \Big(\mathbf{y}_t - \overline{\mathbf{y}} - \sum_{j=1}^p \boldsymbol{\Phi}_j(\mathbf{y}_{t-j} - \overline{\mathbf{y}})\Big) - \mathbf{B}(\tau)\right\| \le$$

$$\le M \cdot \left(\sup_{0\le\tau\le 1}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor T\tau\rfloor}\big(\mathbf{y}_t - \overline{\mathbf{y}}\big) - \boldsymbol{\Sigma}\mathbf{B}(\tau)\right\| + \sup_{0\le\tau\le 1}\left\|\frac{1}{\sqrt{T}}\sum_{j=1}^p \boldsymbol{\Phi}_j \sum_{t=1}^{\lfloor T\tau\rfloor}\big(\mathbf{y}_{t-j} - \overline{\mathbf{y}}\big)\right\|\right) =:$$

$$= M\big(A_T + B_T\big),$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Omega}\big(\mathbf{I}_n - \mathbf{K}^\top\big)^{-1}$. As $\mathbf{B}(\tau) = \mathbf{W}(\tau) - \tau\mathbf{W}(1)$, it holds that

$$A_T \le \sup_{0\le\tau\le 1}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor T\tau\rfloor}\big(\mathbf{y}_t - \boldsymbol{\mu}\big) - \boldsymbol{\Sigma}\mathbf{W}(\tau)\right\| + \sup_{0\le\tau\le 1}\left\{|\tau|\cdot\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\big(\mathbf{y}_t - \boldsymbol{\mu}\big) - \boldsymbol{\Sigma}\mathbf{W}(1)\right\|\right\}.$$

Both latter terms are $o_{\mathsf{P}}(1)$ according to Proposition 3 and hence $A_T = o_{\mathsf{P}}(1)$. Term B_T is treated as follows:

$$B_T \le M \sup_{1\le k\le T}\sup_{1\le s\le p}\frac{1}{\sqrt{T}}\left\|\sum_{t=1}^{k-s}\big(\mathbf{y}_t - \boldsymbol{\mu}\big) - \mathbf{W}(k-s)\right\| + M \sup_{1\le k\le T}\frac{k}{T\sqrt{T}}\left\|\sum_{t=1}^{T}\big(\mathbf{y}_t - \boldsymbol{\mu}\big) - \mathbf{W}(T)\right\| +$$

$$+ M\frac{1}{\sqrt{T}}\sup_{1\le k\le T}\sup_{1\le s\le p}\left\|\mathbf{W}(k-s) - \frac{k}{T}\mathbf{W}(T)\right\|$$

which is together $o_{\mathsf{P}}(1)$ due to Proposition 3 and Theorem 2.

The supremum of the norm of the difference between (5) and (7) is bounded by

$$\left\|\left(\mathbf{I}_n-\widehat{\mathbf{K}}\right)^{\top}\widehat{\boldsymbol{\Omega}}^{-1}-\left(\mathbf{I}_n-\mathbf{K}\right)^{\top}\boldsymbol{\Omega}^{-1}\right\|\cdot\left(\sup_{1\le k\le T}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{k}\left(\widehat{\boldsymbol{\varepsilon}}_t-\boldsymbol{\varepsilon}_t\right)\right\|+\sup_{1\le k\le T}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{k}\boldsymbol{\varepsilon}_t\right\|\right)+$$

$$+M\sup_{1\le k\le T}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{k}\left[\left(\widehat{\boldsymbol{\mu}}-\overline{\mathbf{y}}\right)+\sum_{j=1}^{p}\left(\widehat{\boldsymbol{\Phi}}_j-\boldsymbol{\Phi}_j\right)\left(\mathbf{y}_{t-j}-\overline{\mathbf{y}}\right)-\sum_{j=1}^{p}\widehat{\boldsymbol{\Phi}}_j\left(\widehat{\boldsymbol{\mu}}-\overline{\mathbf{y}}\right)\right]\right\|. \tag{8}$$

According to Proposition 4, $\left\|\left(\mathbf{I}_n-\widehat{\mathbf{K}}\right)^{\top}\widehat{\boldsymbol{\Omega}}^{-1}-\left(\mathbf{I}_n-\mathbf{K}\right)^{\top}\boldsymbol{\Omega}^{-1}\right\|=\mathcal{O}\left(\sqrt{T^{-1}\log\log T}\right)$ a.s. Expansion (10) and Proposition 4 yield that $\frac{1}{\sqrt{T}}\sum_{t=1}^{k}\left(\widehat{\boldsymbol{\varepsilon}}_t-\boldsymbol{\varepsilon}_t\right)=\mathcal{O}(\sqrt{\log\log T})$ a.s. Law of the iterated logarithm for i.i.d. random vectors $\boldsymbol{\varepsilon}_t$ gives us $\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{k}\boldsymbol{\varepsilon}_t\right\|=\mathcal{O}(\sqrt{\log\log T})$ a.s. The latest supremum in (8) is $o_{\mathsf{P}}(1)$ due to Proposition 4 and the fact that $\|\widehat{\boldsymbol{\mu}}-\overline{\mathbf{y}}\|=o(T^{\frac{1}{\nu}-1})$ a.s., for some $\nu>2$. This completes the proof for the part (i).

(ii) Let us suppose we want to test changes in $\boldsymbol{\Phi}_s$, $s=1,\dots,p$. Let $\mathbf{v}_t:=\mathsf{vec}\left(\mathbf{y}_{t-1},\dots,\mathbf{y}_{t-p}\right)\in\mathbb{R}^{np}$, $\boldsymbol{\mu}_{\boldsymbol{p}}:=\mathsf{vec}(\boldsymbol{\mu},\dots,\boldsymbol{\mu})\in\mathbb{R}^{np}$, $\mathbf{M}_t:=\left(\mathbf{v}_t-\boldsymbol{\mu}_{\boldsymbol{p}}\right)^{\top}\otimes\mathbf{I}_n\in\mathbb{R}^{n\times n^2p}$ and $\widehat{\mathbf{M}}_t:=\left(\mathbf{v}_t-\widehat{\boldsymbol{\mu}_{\boldsymbol{p}}}\right)^{\top}\otimes\mathbf{I}_n$. Then

$$\left[\frac{\partial}{\partial\boldsymbol{\phi}}\ell_k(\boldsymbol{\theta})\right]_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}=\sum_{t=1}^{k}\widehat{\mathbf{M}}_t^{\top}\widehat{\boldsymbol{\Omega}}^{-1}\left(\mathbf{y}_t-\widehat{\boldsymbol{\mu}}-\widehat{\mathbf{M}}_t\widehat{\boldsymbol{\phi}}\right)=\sum_{t=1}^{k}\widehat{\mathbf{M}}_t^{\top}\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\varepsilon}}_t.$$

Replacing $\widehat{\boldsymbol{\Omega}}$ with $\boldsymbol{\Omega}$ does not change the asymptotic distribution. By the invariance principle for the martingale difference sequence $\left\{\mathbf{M}_t^{\top}\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_t\right\}$ it holds

$$\sup_{0\le\tau\le1}\left\|\frac{1}{\sqrt{T}}\left(\sum_{t=1}^{\lfloor T\tau\rfloor}\left(\mathbf{M}_t^{\top}\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_t\right)-\frac{\lfloor T\tau\rfloor}{T}\sum_{t=1}^{T}\left(\mathbf{M}_t^{\top}\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_t\right)-\boldsymbol{\Sigma}_1\mathbf{B}(\tau)\right)\right\|=o_{\mathsf{P}}(1), \tag{9}$$

for some $\boldsymbol{\Sigma}_1>0$. We now have to show that the error committed by replacing the parameters in the formula (9) with their maximum likelihood parameters is negligible:

$$\left(\sum_{t=1}^{k}\left(\widehat{\mathbf{M}}_t^{\top}\boldsymbol{\Omega}^{-1}\widehat{\boldsymbol{\varepsilon}}_t\right)-\sum_{t=1}^{k}\left(\mathbf{M}_t^{\top}\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_t\right)\right)+\left(-\frac{k}{T}\sum_{t=1}^{T}\left(\widehat{\mathbf{M}}_t^{\top}\boldsymbol{\Omega}^{-1}\widehat{\boldsymbol{\varepsilon}}_t\right)+\frac{k}{T}\sum_{t=1}^{T}\left(\mathbf{M}_t^{\top}\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_t\right)\right)=:\mathbf{R}_{k,T}+\mathbf{S}_{k,T}.$$

Term $\mathbf{R}_{k,T}$ can be treated as follows:

$$\begin{aligned}\mathbf{R}_{k,T}&=\sum_{t=1}^{k}\left[\left((\mathbf{v}_t-\boldsymbol{\mu}_{\boldsymbol{p}}-\widehat{\boldsymbol{\mu}_{\boldsymbol{p}}}+\boldsymbol{\mu}_{\boldsymbol{p}})\otimes\mathbf{I}_n\right)\boldsymbol{\Omega}^{-1}\left(\widehat{\boldsymbol{\varepsilon}}_t-\boldsymbol{\varepsilon}_t+\boldsymbol{\varepsilon}_t\right)\right]-\sum_{t=1}^{k}\left[\left((\mathbf{v}_t-\boldsymbol{\mu}_{\boldsymbol{p}})\otimes\mathbf{I}_n\right)\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_t\right]=\\&=\sum_{t=1}^{k}\left[\left((\mathbf{v}_t-\boldsymbol{\mu}_{\boldsymbol{p}})\otimes\mathbf{I}_n\right)\boldsymbol{\Omega}^{-1}\left(\widehat{\boldsymbol{\varepsilon}}_t-\boldsymbol{\varepsilon}_t\right)\right]+\sum_{t=1}^{k}\left[\left((\boldsymbol{\mu}_{\boldsymbol{p}}-\widehat{\boldsymbol{\mu}_{\boldsymbol{p}}})\otimes\mathbf{I}_n\right)\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_t\right]+\\&+\sum_{t=1}^{k}\left[\left((\boldsymbol{\mu}_{\boldsymbol{p}}-\widehat{\boldsymbol{\mu}_{\boldsymbol{p}}})\otimes\mathbf{I}_n\right)\boldsymbol{\Omega}^{-1}\left(\widehat{\boldsymbol{\varepsilon}}_t-\boldsymbol{\varepsilon}_t\right)\right]=:\mathbf{R}_{k,T}^{(1)}+\mathbf{R}_{k,T}^{(2)}+\mathbf{R}_{k,T}^{(3)}.\end{aligned}$$

By Proposition 4 and the law of the iterated logarithm for $\{\varepsilon_t\}$

$$\|\mathbf{R}_{k,T}^{(2)}\|=\left\|\left((\boldsymbol{\mu}_{\boldsymbol{p}}-\widehat{\boldsymbol{\mu}_{\boldsymbol{p}}})\otimes\mathbf{I}_n\right)\boldsymbol{\Omega}^{-1}\sum_{t=1}^{k}\boldsymbol{\varepsilon}_t\right\|=\mathcal{O}\left(\sqrt{\frac{\log\log T}{T}}\right)\cdot\mathcal{O}(\sqrt{T\log\log T})=\mathcal{O}(\log\log T)\quad\text{a.s.}$$

Before analyzing $\mathbf{R}_{k,T}^{(1)}$ and $\mathbf{R}_{k,T}^{(3)}$ we write

$$\begin{aligned}\widehat{\boldsymbol{\varepsilon}}_t-\boldsymbol{\varepsilon}_t&=\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}}-\sum_{j=1}^{p}\widehat{\boldsymbol{\Phi}}_j\left(\mathbf{y}_{t-j}-\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}}+\boldsymbol{\mu}\right)+\sum_{j=1}^{p}\boldsymbol{\Phi}_j\left(\mathbf{y}_{t-j}-\boldsymbol{\mu}\right)=\\&=\left(\mathbf{I}_n-\mathbf{K}\right)(\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}})+\sum_{j=1}^{p}\left(\boldsymbol{\Phi}_j-\widehat{\boldsymbol{\Phi}}_j\right)\left(\mathbf{y}_{t-j}-\boldsymbol{\mu}\right)+\sum_{j=1}^{p}\left(\widehat{\boldsymbol{\Phi}}_j-\boldsymbol{\Phi}_j\right)\left(\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}\right).\end{aligned} \tag{10}$$

Imposing the latter expansion (10) into $\mathbf{R}_{k,T}^{(1)}$ and $\mathbf{R}_{k,T}^{(3)}$, using Proposition 4 and the invariance principle we obtain that $\left\|\mathbf{R}_{k,T}^{(3)}\right\|=\mathcal{O}(\log\log T)$ a.s. and $\left\|\mathbf{R}_{k,T}^{(1)}\right\|=\mathcal{O}(\log\log T)$ a.s. Term $\mathbf{S}_{k,T}$ can be analyzed in the similar way, because $\mathbf{S}_{k,T}=-\frac{k}{T}\mathbf{R}_{T,T}$. Hence, $\sup_{1\le k\le T}\|\mathbf{R}_{k,T}+\mathbf{S}_{k,T}\|=\mathcal{O}(\log\log T)$ a.s., which concludes the proof of the theorem. □

3 Test and small simulation study

The application of Theorem 1 for detecting abrupt changes in $\mathbb{VAR}(p)$ model is as follows: If one wants to test for changes in d parameters, $d = 1, \ldots, r$, on the overall significance level α, the null hypothesis H_0 is rejected if

$$\max_{1<k<T} \left| \widehat{B}_j\big(k/T\big) \right| \geq C(\alpha^*), \quad \text{for at least one } j, \tag{11}$$

where $C(\alpha^*)$ is a critical value and $\alpha^* = 1 - (1-\alpha)^{\frac{1}{d}}$. If (11) holds we conclude that there is a change in parameter θ_j, $j = 1, \ldots, r$. Since (4) holds, critical value $C(\alpha^*)$ can be obtained from the limiting process:

$$\mathsf{P}\Big[\sup_{0\leq\tau\leq 1} \big|B_1(\tau)\big| > x \Big] = 1 + \sum_{k=-\infty}^{\infty} (-1)^{k+1} \mathrm{e}^{-2k^2x^2},$$

see [1], p. 103. In order to explore accuracy of the approximation, the simulation study was carried out. We considered a bivariate $\mathbb{VAR}(1)$ model of the form $\mathbf{y}_t = \mathbf{\Phi}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$, $t = 1, \ldots, 200$, where $\boldsymbol{\varepsilon}_t$ was the centered i.i.d. Gaussian sequence of random vectors with variance matrix $\mathbf{\Omega}$ and

$$\mathbf{\Phi} = \begin{pmatrix} 0.5 & 0.1 \\ 0.2 & 0.1 \end{pmatrix} \quad \text{and} \quad \mathbf{\Omega} = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}.$$

Figure 1 illustrates a reasonable convergence of the empirical distribution function of $\sup_{0\leq\tau\leq 1} \big|\widehat{B}_j(\tau)\big|$, $j = 1, \ldots, 6$, based on 1 000 simulations of the latter model to the limiting distribution. The accuracy does not get worse too much even if some more persistent autoregressive matrix is used. One obtains satisfactory results also for moderate sample sizes. We do not report any figures to save space.

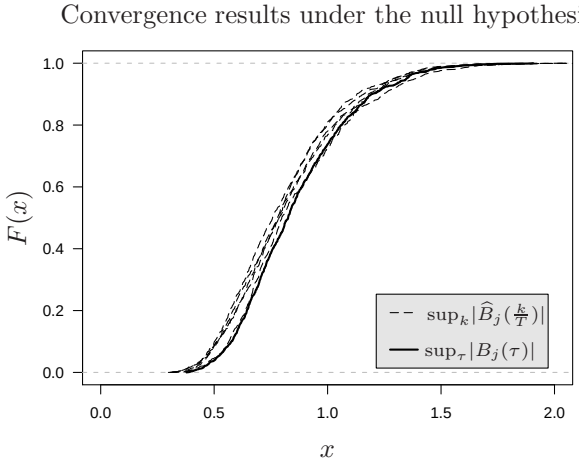


Figure 1: Empirical distribution functions of $\sup_\tau \big|\widehat{B}_j(\tau)\big|$ and the limiting distribution.

4 Conclusion

The article presents the statistical procedure to detect changes in the mean and autoregressive parameters of $\mathbb{VAR}(p)$ models. We have shown that the test statistic based on the efficient score vector converges to the Brownian bridge. Simulation results confirm that the convergence is relatively rapid even for moderate length of time series. Repeated Monte-Carlo experiments for moderate sample sizes reveal that the components of the score vector can be regarded as asymptotically uncorrelated and asymptotically Gaussian based on the p-values of multivariate Shapiro-Wilk test. The extension of the procedure to cover change detection in the variance matrix is also possible and it is the subject of the future study.

A Supplement

The first theorem describes the boundedness of increments of the multivariate Wiener process.

Theorem 2. *If* $\mathbf{W}$ *is a multivariate Brownian motion with independent components, it holds*

$$\limsup_{k\to\infty} \sup_{0\le s\le p} \|\mathbf{W}(k-s)-\mathbf{W}(k)\| = \mathcal{O}\big(\sqrt{\log k}\big) \quad \text{a.s.}$$

Proof. Straightforward using Theorem 1.2.1 of [2]. □

Proposition 3. *Under Assumptions (A.1) – (A.3) it holds that*

$$\sup_{0\le\tau\le 1} \left\| \frac{1}{\sqrt{T}}\Big(\sum_{t=1}^{\lfloor T\tau\rfloor} (\mathbf{y}_t-\boldsymbol{\mu})\Big) - \boldsymbol{\Sigma}\mathbf{W}(\tau)\right\| = o_{\mathsf{P}}(1).$$

Proof. The invariance principle for the stationary linear processes yields

$$\Big\| \sum_{t=1}^{s} (\mathbf{y}_t-\boldsymbol{\mu}) - \boldsymbol{\Sigma}\mathbf{W}(s)\Big\| = o\big(s^{\frac{1}{2}-\delta}\big) \quad \text{a.s., } \delta>0.$$

Let $s := \lfloor T\tau\rfloor$, $0\le\tau\le 1$. Then

$$\frac{1}{\sqrt{T}}\Big\| \sum_{t=1}^{\lfloor T\tau\rfloor} (\mathbf{y}_t-\boldsymbol{\mu}) - \boldsymbol{\Sigma}\mathbf{W}(T\tau)\Big\| = o(1) \quad \text{a.s.} \quad T\to\infty, \tag{12}$$

from which

$$\left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T\tau\rfloor} (\mathbf{y}_t-\boldsymbol{\mu}) - \boldsymbol{\Sigma}\mathbf{W}(\tau)\right\| = \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T\tau\rfloor} (\mathbf{y}_t-\boldsymbol{\mu}) - \frac{1}{\sqrt{T}}\boldsymbol{\Sigma}\mathbf{W}(T\tau) + \frac{1}{\sqrt{T}}\boldsymbol{\Sigma}\mathbf{W}(T\tau) - \boldsymbol{\Sigma}\mathbf{W}(\tau)\right\| \le$$

$$\le \frac{1}{\sqrt{T}}\Big\| \sum_{t=1}^{\lfloor T\tau\rfloor} (\mathbf{y}_t-\boldsymbol{\mu}) - \boldsymbol{\Sigma}\mathbf{W}(T\tau)\Big\| + \|\boldsymbol{\Sigma}\| \cdot \left\| \frac{1}{\sqrt{T}}\mathbf{W}(T\tau) - \mathbf{W}(\tau)\right\|. \tag{13}$$

The first addend in (13) is $o(1)$ a.s. due to (12). For the second addend in (13) we use the well known fact that $\frac{1}{\sqrt{T}}\mathbf{W}(T\tau) \overset{d}{\equiv} \mathbf{W}(\tau)$ from which follows that $\frac{1}{\sqrt{T}}\mathbf{W}(T\tau) - \mathbf{W}(\tau) = o_{\mathsf{P}}(1)$. The result is therefore $o_{\mathsf{P}}(1)$. Because $h(x) = \sup_{0\le t\le 1} x(t)$ is continuous then the statement of the proposition holds. □

Proposition 4. *Under Assumptions (A.1)–(A.3) it holds* $\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}\| = \mathcal{O}\big(\sqrt{T^{-1}\log\log T}\big)$ a.s., $\|\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi}\| = \mathcal{O}\big(\sqrt{T^{-1}\log\log T}\big)$ a.s. *and* $\|\widehat{\boldsymbol{\Omega}}-\boldsymbol{\Omega}\| = \mathcal{O}\big(\sqrt{T^{-1}\log\log T}\big)$ a.s., *as* $T\to\infty$.

Proof. The proof follows from the asymptotic equivalence of ML and least squares estimators and the multivariate extension of Theorem 2.1 in [4]. □

Acknowledgements

The work was supported by the grants SVV 265315/2012 and GAUK 586712/2012.

References

[1] Billingsley P.: *Convergence of Probability Measures*. 2nd ed., Wiley, New York, 1999.

[2] Csörgő M. and Révész P.: *Strong Approximations in Probability and Statistics*. Akadémiai Kiadó, Budapest, 1981.

[3] Gombay E.: Change Detection in Autoregressive Time Series. *Journal of Multivariate Analysis* **99**(3), (2008), 451–464.

[4] Gombay E. and Serban D.: Monitoring Parameter Change in AR(p) Time Series Models. *Journal of Multivariate Analysis* **100**(4), (2009), 715–725.

[5] Qu Z. and Perron P.: Estimating and Testing Structural Changes in Multivariate Regressions. *Econometrica* **75**(2), (2007), 459–502.

[6] Lütkepohl H.: *New Introduction to Multiple Time Series Analysis*. Springer, Berlin, 2005.

ARIMA vs. ARIMAX – which approach is better to analyze and forecast macroeconomic time series?

Ďurka Peter[1], Pastoreková Silvia[2]

Abstract. Nowadays, there are a lot of methods and techniques to analyze and forecast time series. One of the most used is methodology based on autoregressive integrated moving average (ARIMA) model by Box and Jenkins [1]. This method uses historical data of univariate time series to analyze its own trend and forecast future cycle.
Time series are often affected by special events such as legislative activities, policy changes, environmental regulations, and similar events, which we shall refer to as intervention events. You can incorporate one or more time series in a model to predict the value of another series, by using a transfer function. Transfer functions can be used both to model and forecast the response series and to analyze the impact of the intervention.
The general transfer function model employed by the ARIMA procedure was discussed by Box and Tiao [2]. When an ARIMA model includes other time series as input variables, the model is sometimes referred to as an ARIMAX model. Pankratz [4] refers to the ARIMAX model as dynamic regression.
In this article, we use both ARIMA and ARIMAX approaches to analyze and forecast macroeconomic time series and decide whether more complex ARIMAX model brings so much better results than simple ARIMA model.

Keywords: ARIMA, transfer function model, TFM, ARIMAX, gross domestic product per capita, forecast.

JEL Classification: C22, C53
AMS Classification: 91B84

1 ARIMA model

An „AutoRegressive Integrated Moving-Average“ (ARIMA) model belongs to the one of the most used methodology approaches for analyzing time series. This is mostly because of it offers great flexibility in analyzing various time series and because of achieving accurate forecasts, too. Its other advantage is that for analyzing single time series it uses its own historical data.

The ARIMA model methodology was first introduced by Box and Jenkins in 1976 [1], and ARIMA models are often referred to as Box-Jenkins models. This approach analyzes univariate stochastic time series, i. e. error term of this time series. For this to be possible, the analyzed time series must be stationary. This means that the mean, variance and covariance of the series are all constant over time. However, most economic and financial time series show trends over time. Stationarity is important because, if the series is non-stationary, all the typical results of the classical regression analysis are not valid. Regressions with non-stationary series may have no meaning and are therefore called „spurious“. Long-term forecasts of a stationary series will converge to the unconditional mean of the series.

ARIMA model (with seasonal terms) can be written as follows:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \Phi_1 y_{t-s} + \Phi_2 y_{t-2s} + ... + \Phi_P y_{t-Ps} + \\ + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - ... - \theta_q a_{t-q} - \Theta_1 a_{t-s} - \Theta_2 a_{t-2s} - ... - \Theta_Q a_{t-Qs} \quad (1)$$

Using backshift (lag) operator we can rewrite (1):

$$\phi_p(B)\Phi_P(B^s) z_t = \theta_q(B)\Theta_Q(B^s) a_t \quad (2)$$

where:
$z_t = (1-B)^d (1-B^s)^D ln(y_t)$
$\phi_p(B)$ – nonseasonal operator of autoregressive process $AR(p)$

1 University of Economics, Department of statistics, Dolnozemská cesta 1, 852 35 Bratislava 5, Slovakia e-mail: peter.durka@euba.sk
2 University of Economics, Department of statistics, Dolnozemská cesta 1, 852 35 Bratislava 5, Slovakia e-mail: pastorekova.silvia@gmail.com

$\theta_q(B)$ – nonseasonal operator of moving average *MA(q)*
$\Phi_P(B^s)$ – seasonal operator of autoregressive process *AR(P)*
$\Theta_Q(B^s)$ – seasonal operator of moving average *MA(Q)*
a_t – error term (white noise)
s – orders of season ($B^s y_t = y_{t-s}$)
d, D – nonseasonal and seasonal orders of differencing (integration)

Then, using more parsimonious notation, we can rewrite (2) as follows:

$$\text{ARIMA } (p, d, q)(P, D, Q)_s, \tag{3}$$

where: p, P – number of autoregressive parameters
q, Q – number of moving average parameters

The Box-Jenkins approach is iterative three-stage modeling approach – identification, estimation and diagnostic checking, and finally forecasting.

In the identification stage, the researcher visually examines the time plot of the series autocorrelation function (ACF) and partial autocorrelation function (PACF). Plotting each observation of the series against time t provides useful information concerning outliers, missing values and structural breaks in the data. The analyzed time series must be stationary. Once stationarity has been achieved (logarithm and/or differences), the next step is to identify the parameters of the model, i. e. AR and MA orders examining ACF and PACF.

In the estimation stage, each of the tentative models is estimated and the various coefficients are examined. The estimated models are compared using the Akaike information criterion and the Schwarz Bayesian criterion and model with the smallest criterion is chosen to get the parsimonious model. The main approaches to fitting Box–Jenkins models are non-linear least squares and maximum likelihood estimation.

In the diagnostic checking stage, the goodness of fit of the model is examined. Residuals should meet white noise assumptions, i. e. autocorrelation, homoskedasticity and normality is tested. If these assumptions are not satisfied, one needs to fit a more appropriate model. Care must be taken here to avoid overfitting.

The main function of ARIMA models is forecasting. Their forecasting ability can be considered when compared to actual time series.

2 ARIMAX – transfer function model

Assume two time series denoted Y_t and X_t, which are both stationary. Then, the *transfer function model* (TFM) can be written as follows:

$$Y_t = C + \nu(B)X_t + N_t \tag{4}$$

where:
Y_t is the output series (dependent variable),
X_t is the input series (independent variable),
C is constant term,
N_t is the stochastic disturbance, i.e. the noise series of the system that is independent of the input series.
$\nu(B)X_t$ is the transfer function (or impulse response function), which allows X to influence Y via a distributed lag.
B is backshift operator, thus we can write

$$\nu(B)X_t = (\nu_0 + \nu_1 B + \nu_2 B^2 + \ldots)X_t, \tag{5}$$

When X_t and N_t are assumed to follow ARMA model, equation (4) is known as the ARMAX model. This ARMAX model is quite different from ARMA model, because we work with two different series X_t and Y_t - output series Y_t is related to input series X_t.

Coefficients ν_j are called impulse response weights, which could be positive or negative. The larger the absolute value of any weight ν_j is, the larger is the response of Y_t to a change in X_{t-j}. Output series might not react immediately to a change in input series, thus some initial ν weights may be equal to zero. The number of ν weights equal to zero is called dead time and is denoted as b (Rublikova, Marek [5]).

Theoretically, the transfer function $\nu(B)X_t$ has an infinite number of coefficients. Then, we can write transfer function as the rational polynomial distributed lag model of finite order as the ratio of a low order polynomials in B:

$$\nu(B)X_t = \frac{\omega_h(B)B^b}{\delta_r(B)}X_t \tag{6}$$

where $\omega_h(B) = \omega_0 + \omega_1 B + \ldots + \omega_h B^h$; $\delta_r(B) = 1 - \delta_1 B - \ldots - \delta_r B^r$; h is the number of terms plus one of the independent variable included; r is the number of terms of the dependent variable included and b is dead time mentioned above already.

Disturbance series N_t can be written in the form of an autoregressive integrated moving average model as follows:

$$N_t = \frac{\theta(B)\Theta(B^s)}{\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D} a_t \tag{7}$$

where a_t is zero mean and normally distributed white noise.

Then, substitute (5) with maximum lag denoted by K (free-form distributed lag model) and (7) into (4), we have transfer function model in its full formula:

$$Y_t = C + \nu_0 X_t + \nu_1 X_{t-1} + \nu_2 X_{t-2} + \cdots + \nu_K X_{t-K} + \frac{\theta(B)\Theta(B^s)}{\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D} a_t \tag{8}$$

Construction of TFM is similar iterative process as construction of univariate Box-Jenkins ARIMA model, i.e. identification, estimation and diagnostic chcecking. After checking there is no feedback from earlier values of the output to current values of the input, we can start with the linear transfer identification method (LTF) to find out the orders (*b, r, h*) of a rational form transfer function (Pankratz [4]). First, we specify free-form distributed lag model in which K is chosen according to the analyst judgment and then we specify low order for disturbance series N_t. Nonlinear least square method can be used to estimate parameters. After estimation of the model, we have to check estimated disturbance series for stationarity by means of sample autocorrelation function and sample partial autocorrelation function. If the disturbance series is not stationary, then it is necessary to difference input and output accordingly. If the disturbance is stationary, then we are going to the stage 2 where we may use preliminary estimated impulse response weights to choose the orders (*b, r, h*) of one/few tentative rational form transfer function(s) to represent $\nu(B)$. We can identify the orders (*b, r, h*) by visually comparing the estimated impulse response function with some common theoretical functions. If the linear transfer function model is adequate then we can compute forecasts. There are several diagnostic checks to decide whether the model is adequate based on the residuals which should be independent as well as input series, e.g. cross-correlation check and/or autocorrelation check.

It is good practice to build an ARIMA model for both the output and the input series before attempting to build a transfer function model (Rublikova, Marek [5])

3 Transfer function model for gross domestic product per capita

In this applied part, we are going to build a transfer function model (TFM) for gross domestic product per capita and unemployment rate. We assume that a change in unemployment rate will affect trend in gross domestic product per capita which will lead to a significant change.

3.1 ARIMA model

First, we are going to find best fitted ARIMA model for output and input series. Output series is gross domestic product per capita (GDPpc). Analyzed quarterly data cover the period from 2000 up to 2011 in thousand EUR constant prices. The source of data is Slovstat database. The plotted series is shown at Fig. 1. Because the series is nonstationary, we use difference and seasonally difference to reach the stationarity. After study of ACF and PACF of stationary series, we identify the right model as ARIMA (0,1,0)(1,1,0) written as (standard error is in parentheses):

$$\underset{(0.1441)}{(1-B)(1-B^4)(1+0.3057B^4)}HDPpc_t = a_t \tag{9}$$

Unemployment rate (UR) is used as input series. It covers quarterly data of period from 2000 up to 2011 in percents. The source of data is Slovstat database, too. The plotted series UR is shown at Fig. 2. The series is nonstationary and just difference is appropriate to transform to its stationarity. Then, according to its ACF and PACF we identify the fitted model as ARIMA (0,1,1)(2,0,0):

$$\underset{(0.1397)\quad (0.1381)}{(1-B)(1-0.5894B^4-0.3545B^8)}UR_t = \underset{(0.1116)}{(1+0.7235B)}a_t \tag{10}$$

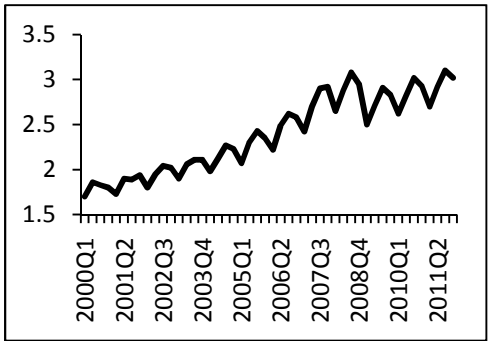


Fig. 1 Gross domestic product per capita in 1000 EUR constant prices, 2000Q1 – 2011Q4
(Source: Slovstat database)

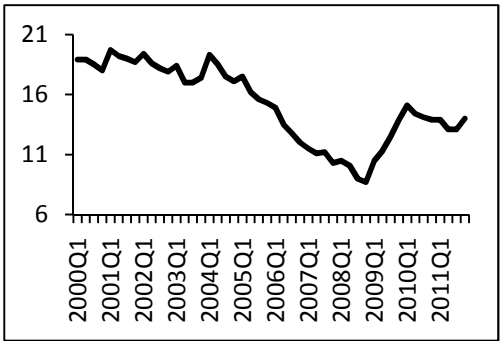


Fig. 2 Unemployment rate in %, 2000Q1 – 2011Q4
(Source: Slovstat database)

3.2 Transfer function model (ARIMAX model)

Now, we apply earlier described LTF method to build transfer function model describing relationship between GDP per capita and unemployment rate. First step is to estimate free-form distributed lag model with input and output series not differenced. We assume maximum lag $K=8$ and noise series N_t is approximated by AR(1). Estimated noise model $\hat{n}_t = y_t - \hat{y}_t$ is stationary. Only statistically significant weight was ν_0. Therefore, the transfer function model has the simplest form with parameters (0, 0, 0) and transfer function is $\nu(B) = \omega_0$.

Next step is to build ARIMA model for noise series N_t. After long identification and estimation, we choose the model in the form ARIMA (0,0,1)(0,0,1):

$$N_t = (1-\theta_1 B)(1-\Theta_1 B^4)a_t \tag{11}$$

Now, we can estimate transfer function model in the form

$$Y_t = C + \nu_0 X_t + (1-\theta_1 B)(1-\Theta_1 B^4)a_t \tag{12}$$

$$GDPpc_t = 3.8983 - 0.0975 UR_t + (1-0.7756B)(1-0.6049B^4)a_t \tag{13}$$
$$(0.2144) \quad (0.0139) \qquad (0.1000) \quad (0.1361)$$

All parameters are statistically significant. Residuals meet white noise assumptions. To check the adequacy of the fitted transfer function model, we calculate the cross-correlation function between TFM residuals and ARIMA model for unemployment rate residuals (estimated above). No values of the cross-correlation function are statistically significant, therefore residuals are not autocorrelated and fitted transfer function model is adequate. The R-Squared statistic indicates that the model as fitted explains 92.7% of the variability in gross domestic product per capita.

3.3 Forecasts of gross domestic product per capita by ARIMA and ARIMAX model

Now, we will compute not only forecasts by fitted transfer function model (ARIMAX) above but also the forecasts for individual series of GDP per capita given by the fitted ARIMA model to compare accuracy of both methods. To compute the TFM forecasts, we need to know the data of input variable unemployment rate in the time of forecasts which we can calculate using ARIMA model methodology.

Quarterly forecasts for 2012 calculated by both methods are in Table 1. Both fitted models assume slight growth of GDP per capita in quarters of 2012. We can see that forecasted data by ARIMAX model are a little lower than simple univariate ARIMA model data.

Model	2012Q1	2012Q2	2012Q3	2012Q4
ARIMA	2.7961	3.0100	3.1961	3.1131
ARIMAX	2.6171	2.5531	2.6824	2.7212

Tab. 1 Forecasts of gross domestic product per capita by ARIMA and ARIMAX in 2012 (in thousand EUR constant prices)

(Source: Authors)

Graphical comparison of raw data of GDP per capita and fitted data and forecasts by both methodologies are at Fig. 3.

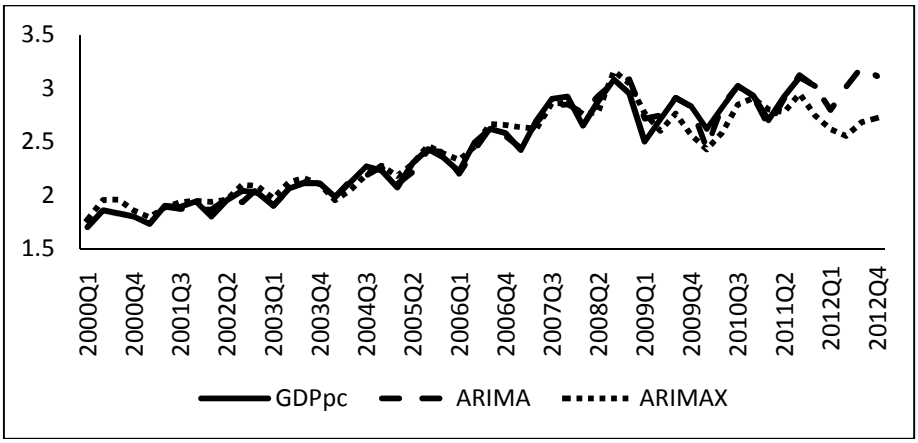


Fig. 3 Raw data of GDP per capita, ARIMA fitted data, ARIMAX fitted data and forecasts (Source: Authors)

As we can see, ARIMA model fits the trend of GDP per capita slightly better than ARIMAX model. ARIMA model mean absolute percentage error is 1.77 % and ARIMAX is 3.78 %, and root mean square error is 0.0653 respectively 0.1162.

4 Conclusions

In this article, we built the transfer function model (ARIMAX) for gross domestic product per capita as an output series and unemployment rate as an input series. Fitted model was adequate and residuals were white noise. The R-Squared statistic indicated that the model as fitted explains 92.7% of the variability in gross domestic product per capita. Next, we calculated quarterly forecasts for 2012 and compared them with forecasts calculated by simple univariate ARIMA model for GDP per capita. Forecasts were slightly different and both assumed growth of GDP per capita. ARIMA model mean absolute percentage error and root mean square error were lower than ARIMAX. ARIMA model seems to be a little accurate than ARIMAX.

Acknowledgement

This paper was elaborated in the framework of the project number IGPM 2317123: Measuring and analyzing of indicators of sustainable development using quantitative methods.

References

[1] Box, G. E. P. – Jenkins, G. M. – Reinsel, G. C.: *Time Series Analysis: Forecasting and Control*. John Wiley & Sons Inc., New York, 2008.

[2] Box, G. E. P. – Tiao, G. C.: Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association.* 1975, vol. 70, no. 349, p. 70-79.

[3] Marek, L.: Transfer function models. *Acta Oeconomica Pragnesia*. 2000, vol. 8, no. 3, p. 83-94

[4] Pankratz, A.: *Forecasting with Dynamic Regression Models*. Wiley-Interscience, 1991.

[5] Rublíková, E. – Marek, L.: Linear transfer function model for outflow rates. *Ekonomické rozhľady.* 2001, vol. 30, no. 4, p. 457-466

Multiple messenger problem

Jan Fábry[1], Maria Kobzareva[2]

Abstract. Multiple messenger problem is a generalization of the problem, where more messengers can be used in the solution. This problem is frequently solved problem in logistic management and it is more appropriate for real-life applications, because companies often use several messengers to transport shipments. The paper describes static problems with multiple messengers and suggests a possibility of solving such problems with modified heuristic methods. To solve multiple messenger problem, modified nearest neighbor heuristics and modified insertion heuristics are used. To reduce transportation costs, the generated solution is improved using modified exchange heuristic method. The main contribution of the paper are applications, developed in VBA in MS Excel that can solve static problems with multiple messengers and that can be beneficial for real companies. Results of the computation experiments are presented in the paper.

Keywords: travelling salesman problem, distribution problem, static multiple messenger problem, modified nearest neighbor heuristics, modified insertion heuristics, modified exchange heuristics

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Messenger problem is a special instance of the travelling salesman problem (TSP), where request consists of a pick-up location and a destination location. It is analogous to dial-a-ride problem described in literature. Psaraftis [8] created exact algorithm for the static problem with one vehicle. Cordeau [2] offers branch-and-cut algorithm for solving this problem. Heuristic algorithms for the transportation of handicapped persons are presented by Toth and Vigo [9]. Jorgensen et al. [7] use genetic algorithms for solving the problem.

Multiple messenger problem, where more than one messenger can be used to transport shipments, consists of generating a set of routes for m messengers. The problem is based on multiple travelling salesman problem [1]. Gavish and Srikanth [5] offer the optimal solution method for large-scale multiple TSP. In multiple messenger problem, all messengers start at and turn back to a depot respecting the essential constraint that a messenger cannot visit the delivery location before the pick-up location. Solution of multiple vehicle dial-a-ride problem using tabu search heuristics is described by Cordeau and Laporte [3]. The purpose of this paper is to present modified heuristic methods, being developed specially for multiple messenger problem. The text consists of the following parts: section 2 describes nearest neighbor heuristic algorithms, section 3 presents insertion heuristic algorithm and section 4 offers exchange heuristic algorithm for solving static multiple messenger problem. In section 5, results of computational experiments are presented and in section 6 some concluding statements are given.

2 Modified nearest neighbor algorithm

Nearest neighbor heuristics for TSP is a sequential construction process that starts by initializing the current route at the depot. Then, the customer nearest to the depot, in terms of a given cost or distance, is selected to extend the route. Subsequently, the customer nearest to the last customer will be added to the route, if its selection does not cause any constraint violation [6]. The process is repeated until the Hamiltonian cycle is generated. In [4], the messenger problem is defined for the network in which all even nodes correspond to

[1] Jan Fábry, University of Economics Prague, Faculty of Informatics and Statistics, Department of Econometrics, Prague 3, W. Churchill sq.4, fabry@ vse.cz

[2] Maria Kobzareva, University of Economics Prague, Faculty of Informatics and Statistics, Department of Econometrics, Prague 3, W. Churchill sq.4, Ph.D. student, mariakobzareva@gmail.com

pick-up locations and all odd nodes to delivery locations. In addition, if a shipment is picked-up in node i, it will be delivered to node $i+1$. The modification of the heuristic method offers in each step such nodes that can be selected for extension of the route. Thus, in the first step, only pick-up locations (even nodes) can be considered. The modified nearest neighbor heuristic algorithm for static multiple messenger problem is presented below. Let denote 1 as a depot and n as a number of all locations (respecting the previous assumption, n must be the odd number).

Algorithm 1

Let T be the maximum number of messengers that can be used in the solution, A_t is a set of locations that had not been visited yet and could be visited by a messenger t at a given step, i.e. at the beginning of the algorithm, all sets A_t contain only even numbers 2, 4, …, $n-1$. S_t is a sequence of locations being subsequently visited by messenger t and i_t is the last node, visited by messenger t, c_{ij} is the minimal distance between nodes i and j, z_t is a distance, which messenger t travels on the route, z is the total distance travelled by all messengers.

Step 1:

$A_t = \{2, 4, \ldots, n\text{-}1\}$, $S_t = \{1\}$, $i_t = 1$, $z_t = 0$, $t = 1, 2, ..., T$.

Step 2:

$$c_{i_m k} = \min_{t=1,2,\ldots,T} \min_{j \in A_t} c_{i_t j},$$

$S_m = S_m + \{k\}$, $z_m = z_m + c_{i_m k}$,

$A_t = A_t - \{k\}$, $t = 1, 2, ..., T$,

if k is even, then $A_m = A_m + \{k+1\}$,

$i_m = k$.

Step 3:

If $A_t = \emptyset$ for all $t = 1, 2, \ldots, T$, go to step 4, else go to step 2.

Step 4:

$S_t = S_t + \{1\}$, $z_t = z_t + c_{i_t 1}$, $t = 1,2, ..., T$ (only for $S_t \neq \{1\}$),

$$z = \sum_{t=1}^{T} z_t .$$

End

3 Modified insertion algorithm

Insertion heuristics is based on extending the current route by inserting one location at each iteration. In the first step, we find the farthest location from the depot for each messenger, and then we insert other nodes trying to achieve minimum extension of current routes. The modification of the insertion algorithm, used for multiple TSP, has to respect the following conditions:

a) related pick-up and delivery locations (shipment) must not be included in two different routes,

b) delivery location for each shipment has to be visited after its picking-up.

The modified insertion heuristic algorithm for multiple messenger problem is presented below. Ordering of all nodes follows the assumptions of the nearest neighbor algorithm presented above.

Algorithm 2

Let T be the number of messengers that will be used to transport all shipments, A_t is a set of locations that had not been visited yet and could be visited by a messenger t at a given step, i.e. at the beginning of the algorithm, all sets A_t contain numbers 2,3, …, n. S_t is a sequence of locations being subsequently visited by messenger t, s_i^t is the element of sequence S_t on position i, h_t is the number of elements in sequence S_t, c_{ij} is the minimal distance between nodes i and j, z_t is a distance, which messenger t travels on the route, z is total distance travelled by all messengers, Δz is an increase in total distance.

Step 1:

$A_t = \{2,3,\ldots, n\}$, $t = 1, 2, \ldots, T$,

for $t = 1, 2, \ldots, T$ repeat:

$c_{1k} = \max_{j \in A_t} c_{1j}$,

$S_t = \{1, k, 1\}$, $z_t = c_{1k} + c_{k1}$, $h_t = 3$; $A_t = A_t - \{k\}$,

for $i = 1,2, ..., T$, $i \neq t$ do:

$A_i = A_i - \{k\}$,

if k is even, then $A_i = A_i - \{k+1\}$,

if k is odd, then $A_i = A_i - \{k-1\}$.

Step 2:

$\Delta z = \infty$,

for $t = 1, 2, \ldots, T$ repeat:

for each $r \in A_t$ repeat

a) if r is even and sequence S_t contains node $r + 1$ on position f, then

$$\Delta z_{rj} = \min_{i=1,2,\ldots,f-1} (c_{s_i^t r} + c_{r s_{i+1}^t} - c_{s_i^t s_{i+1}^t}),$$

b) if r is odd and sequence S_t contains node $r - 1$ on position f, then

$$\Delta z_{rj} = \min_{i=f,f+1,\ldots,h_t-1} (c_{s_i^t r} + c_{r s_{i+1}^t} - c_{s_i^t s_{i+1}^t}),$$

c) else

$$\Delta z_{rj} = \min_{i=1,2,\ldots,h_t-1} (c_{s_i^t r} + c_{r s_{i+1}^t} - c_{s_i^t s_{i+1}^t}).$$

If $\Delta z_{rj} < \Delta z$, then $\Delta z = \Delta z_{rj}$, $v = r$, $q = j$, $m = t$.

Insert node v to sequence S_m after element s_q^m,

$h_m = h_m + 1$,

$z_m = z_m + \Delta z$,

$A_m = A_m - \{v\}$,

for $i = 1,2, ..., T$, $i \neq m$ do:

$A_i = A_i - \{v\}$,

if v is even, then $A_i = A_i - \{v+1\}$,

if v is odd, then $A_i = A_i - \{v-1\}$.

Step 3:

If $A_t = \emptyset$, for all $t = 1, 2, \ldots, T$, then go to step 4, else go to step 2.

Step 4:

$$z = \sum_{t=1}^{T} z_t .$$

End

4 Modified exchange algorithm

Both nearest neighbor algorithm and insertion algorithm described above generate the solution of multiple messenger problem. It can be improved using modified exchange algorithm. The main idea is to exclude shipments from generated routes and include them in other routes if this exchange is advantageous.

Algorithm 3

Let T be the number of messengers with routes generated by Algorithm 1 or Algorithm 2, S_t is a sequence of locations to be subsequently visited by messenger t, s_i^t is the element of sequence S_t on position i.

Step 1:

For each messenger route, find a shipment, i.e. pair of nodes i (even) and $i + 1$ (odd), which will bring maximum reduction of the length of the route when excluded from the route. These shipments are candidates for possible exchange. Denote the reductions as Δzk_t ($t = 1, 2, \ldots, T$). Select route m with the maximal reduction:

$$\Delta zk_m = \max_{t=1,2,\ldots,T} \Delta zk_t.$$

Let this reduction correspond to the extraction of nodes r and $r + 1$ from route m.

Step 2:

For $t = 1, 2, \ldots, T$, $t \neq m$ determine the value of minimum extension of the appropriate route when inserting nodes r and $r + 1$. Denote these extensions as Δpr_t. Select route u with the minimal extension:

$$\Delta pr_u = \min_{\substack{t=1,2,\ldots,T \\ t \neq m}} \Delta pr_t.$$

Let the extension of route u correspond to visiting node r after node s_k^u, and node $r + 1$ after node s_l^u, eventually visiting both nodes after node s_k^u [4].

Step 3:

If $\Delta zk_m > \Delta pr_u$, then accomplish the exchange and continue with step 1, else go to the end.

End

The algorithm can be modified in several ways. For example, if at the step 3 the algorithm stopped because of no effective exchange was found, we could continue with step 1 and try to reduce the length of route with the second greatest value of Δzk_t, eventually the third one, etc. More time demanding modification is based on comparison of maximum reductions of all routes with minimum extensions of all other routes after the inclusion of corresponding pairs of nodes. Thus, in step 3, there is accomplished the exchange corresponding to the maximum positive difference $\Delta zk_i - \Delta pr_j$ ($i,j = 1,2, \ldots, T$; $i \neq j$). Another opportunity how to use effectively the modified exchange algorithm is to try to reduce the length of the longest route regardless this reduction is maximum within all routes, eventually to forbid the assignment of additional shipments to busy messengers with long routes.

5 Computational experiments

To perform computational experiments, we generated 25 multiple messenger problems with 21 nodes. Modified nearest neighbor heuristics and insertion heuristics were applied to solve all generated instances. Then we applied modified exchange heuristics to improve obtained results. The instances were divided into 6 groups; each group had different variation of minimal distances between nodes. In the first group, minimal distances between nodes varied up to 60 km, in the last group to 200 km. Each group had 5 generated cases. Figure1 contains the results of computational experiments with the application of the algorithms generating multiple routes. Figure 2 illustrates results of applying modified exchange heuristics on the generated routes. Figures 3 and 4 present the efficiency of the use of exchange algorithm, applied on results, obtained by modified nearest neighbor heuristics and insertion heuristics.

It is obvious that in given set of instances nearest neighbor heuristics offers better solution than insertion heuristics in 96 % of cases, which could be partly given by the structure of minimal distances matrix. Nearest neighbor heuristics turns to be very efficient, as in every step it works only with nodes that can be inserted after the "last" node, so we do not have to test, whether we can add the current node. For example, at the beginning

only pick-up (even) nodes can be selected, because no shipments are processed currently, i.e. there is nothing to deliver.

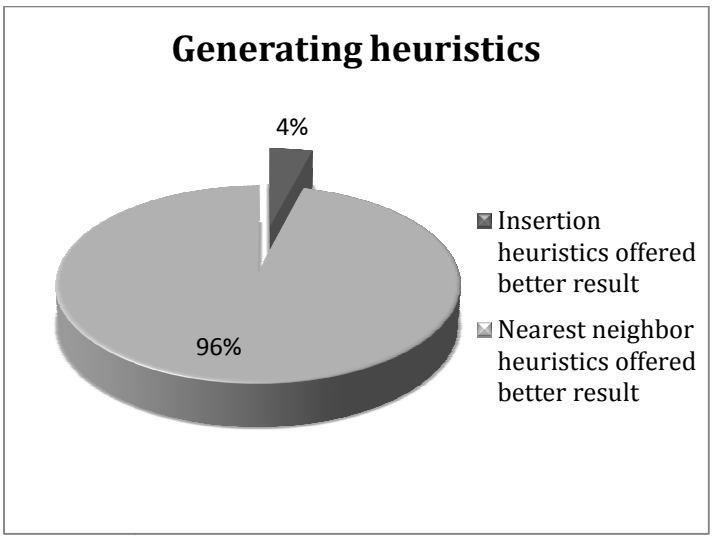


Figure 1 Comparison of modified heuristics for generating solution

When the modified exchange heuristics was applied on solutions generated by nearest neighbor heuristics and insertion heuristics, 63 % of results were better for nearest neighbor algorithm, while 37 % of them were better for insertion algorithm. It supports the idea that the modification of nearest neighbor heuristics for multiple messenger problem was designed successfully. We can also consider that insertion heuristics can achieve better solution in almost 40 % of cases after applying modified exchange heuristics on results, thus combination of insertion and exchange methods can be beneficial.

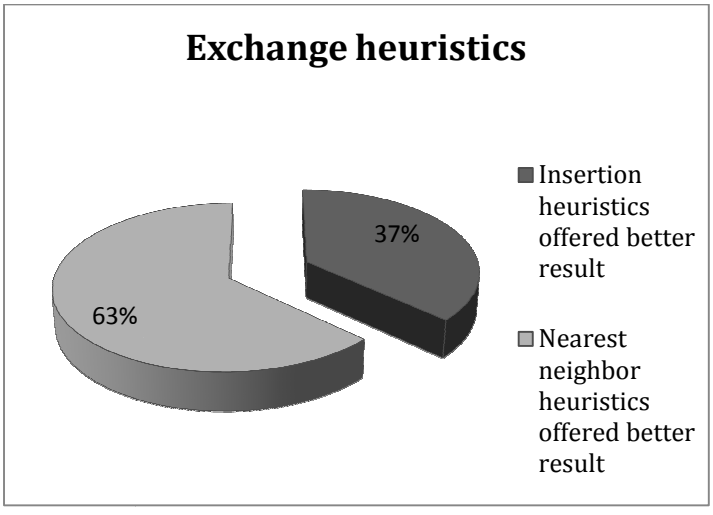


Figure 2 Modified exchange heuristics results

After applying exchange heuristics on solution, generated by modified nearest neighbor heuristics, the results (in terms of total travelled distance) were improved by 3 to 37 %. This fact suggests the idea that modified exchange heuristics is efficient and should be used to improve the solution, obtained by other heuristic methods. In 12 cases out of 25, the improvement after using exchange heuristics was greater than 15 %, only in 3 cases out of 25, the improvement was less than 5 %. In 2 cases, exchange heuristics was unable to find better solution. The average efficiency of solution improvement by modified exchange heuristics is 16 %.

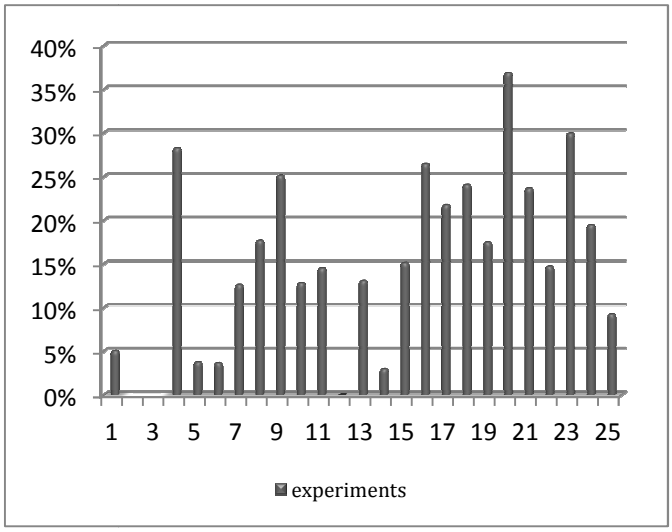


Figure 3 Efficiency of modified exchange heuristics applied on nearest neighbor heuristics

After applying exchange heuristics on solution generated by modified insertion heuristics, the results were improved by 9 to 54 %. We can see that modified exchange heuristics is even more effective when applied on the results, obtained by insertion heuristics. In 19 cases out of 25, the improvement after using exchange heuristics was greater than 20 %, only in 1 case out of 25, the improvement was less than 10 %. In 12 cases out of 25, the improvement was greater than 30 %; exchange heuristics was able to find better solution in every case. The average effectiveness of modified exchange heuristics is 29 %. This fact supports the idea that modified exchange heuristics is efficient and should be used to improve the solution, obtained by different heuristic methods, especially the insertion algorithm.

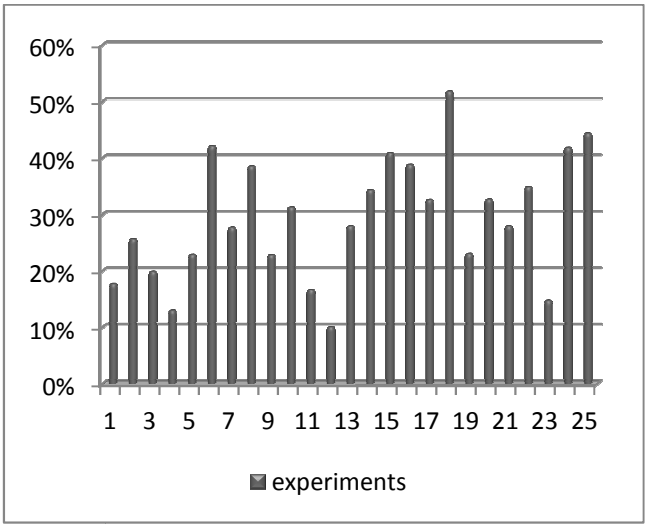


Figure 4 Efficiency of modified exchange heuristic applied on insertion heuristics

6 Conclusions

This paper offers modifications of nearest neighbor algorithm and insertion algorithm to solve multiple messenger problem. As the result, multiple routes are generated. For improvement of solutions, we modified the exchange algorithm, which enables to exclude shipments from some routes and include them in other routes to minimize total distance travelled by all messengers. Beside detailed description of all algorithms, we also present results of computational experiments to determine efficiency of proposed methods.

Acknowledgements

This publication was supported by the project F4/18/2011 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*.Vol. 34, 2006, pp. 209-219.

[2] Cordeau, J.-F. A Branch-and-Cut Algorithm for the Dial-a-Ride Problem. *Operations Research*, Vol. 54, 2006, pp. 573-586.

[3] Cordeau, J.-F. and Laporte, G. A Tabu Search Heuristics for the Static Multi-Vehicle Dial-a-Ride Problem. *Transportation Research Part B*. Vol. 37, 2003, pp. 579-594.

[4] Fábry, J. Dynamic Messenger Problem. *Communications*, Vol. 9 (4), 2007, pp. 66-69.

[5] Gavish, B. and Srikanth, K. An Optimal Solution Method for Large-Scale Multiple Traveling Salesman Problems. *Operations Research*, Vol. 34, 1986, pp. 698-717.

[6] Hosny, M. Investigating heuristic and meta-heuristic algorithms for solving pickup and delivery problems, dissertation, *Cardiff University*, 2010.

[7] Jorgensen, R.M., Larsen, J., Bergvinsdottir, K.B. Solving the Dial-a-Ride Problem using Genetic Algorithms. *Journal of the Operational Research Society*, Vol. 58 (10), 2006, pp. 1321-1331.

[8] Psaraftis, H.N. An Exact Algorithm for the Single Vehicle Many-to-Many Dial-a-Ride Problem with Time Windows. Transportation Science., Vol. 17, 1983, pp. 351-357.

[9] Toth, P., Vigo, D. Heuristic Algorithms for the Handicapped Persons Transportation Problem. Transportation Science 31, 1997, pp. 60-71.

Microeconomic analysis of cartel equilibrium optimization model

Eleonora Fendeková [1], Michal Fendek [2]

Abstract. Cartel as a market structure represents a specific form of oligopoly where an agreement is made between legally independent economic subjects in order to restrict the mechanism of economic competition.
Existence of the cartels is in sharp contrast with generally accepted principles and practices of economic competition protection. In developed economies of the EU and the world the governmental institutions are established to control and guarantee the conditions of competition.
In this paper we will present the mathematical formalization of the model of equilibrium price and supply of the cartel agreement participants and point out the social inefficiency of such decision-making scheme. In more detail we will discuss the analysis of the price cartel considering various costs of the producers. We will study the properties of the cartel profit optimization problem considering various marginal costs of homogeneous production of the cartel participants and point out the interpretation possibilities of solving this optimization problem. We will also point out some interesting economically interpretable implications of the Kuhn-Tucker optimality conditions in optimal decision-making of the cartel subjects in the context of its behavior on the market of imperfect competition.ext goes here.

Keywords: oligopoly market, equilibrium model, Lagrange function, Kuhn–Tucker optimality conditions, market price of cartel.

JEL Classification: L11, L13, D42, D43,
AMS Classification: 90C30, 90C46

1 Introduction

A cartel is generally perceived as a specific market form of oligopoly where cartel subjects accede to a discreet agreement between formally legally independent economic subjects which together enter into a contract with an aim of reaching a more favorable position on relevant market and thus eliminate the mechanism of competition. Cartels which, based on the agreed market strategies, may follow common price strategy, set their production quotes or divide the market are forbidden in the European Union countries as well as in many countries outside the EU.

A cartel is a specific case of oligopoly with an unspecified number of buyers but only a small number of sellers. The upper limit of the number of the sellers in the market structure for it to be defined as oligopoly is not explicitly defined. Pepall [6] shows, that the key issue is not the number of the sellers but the way they communicate with each other, how they react to their individual intentions and how they jointly address the conditions and attributes of the oligopoly market equilibrium, that is how they resolve the question of market price of their products, how they set the total supply of the sector and how they agree on the individual contribution of the oligopoly subjects to the creation of total oligopoly supply on the relevant market.

It should be noted that each action of a particular firm in an oligopoly affects the behavior of other firms on the market. Price lowering of one firm will likely decrease a market share of other firms on the production of a sector. In other words, responses of the competitors in oligopoly may have a significant effect on a result of managerial decision making on an oligopoly market. It is therefore clear that a decision making about optimal price and supply in an oligopoly is far more complicated than in other market structures.

[1] University of Economics Bratislava, Department of Business Economics, Dolnozemska cesta 1/b, 852 35 Bratslava, Slovakia, e-mail: nfendek@dec.euba.sk.
[2] University of Economics Bratislava, Department of Operations Research and Econometrics, Dolnozemska cesta 1/b, 852 35 Bratislava, Slovakia, e-mail: fendek@dec.euba.sk

In some situations the subjects may simply ignore most of the actions of their competitors, in other cases, however, a price war between the oligopoly subjects may strike as a reaction to seemingly innocent price change. Multiple factors such as maturity of a sector, nature of production and also business methods can determine a way in which firms respond to competitors' behavior.

Methodological problems of formulating oligopoly models rise from the great diversity of ways in which firms can interact and conclude agreements on the distribution, market shares and market prices. Simply said, there is no general model of oligopoly.

In the paper we present the results of microeconomic analysis of a cartel equilibrium model and show how a method of setting an optimal price and supply of a cartel limits the laws of competition. We study the properties of a cartel profit optimization problem considering various marginal costs of homogeneous production of the cartel subjects and point out some interesting, economically interpretable implications of Kuhn-Tucker optimality conditions in a cartel subject's optimal decision making problem in a context of its behavior on the market of imperfect competition.

2 Model of Supply and Price Equilibrium in a Cartel

Let's study a situation in a sector where we assume the individual producers agree on a common course of action while setting a volume of their production, market shares and market price. So the producers will de facto act as a monopoly even though formally they remain independent firms.

If the producers ignored a behavior of the other producers while making their decisions, every firm would independently solved its own profit maximization problem based on classical optimality conditions.

Carlton [1] shows, that if, however, the firms won't proceed independently but agree on common approach to production and sales, by which they violate the conditions of competition, their challenge will then be to find the answers to following questions:

- What will be the equilibrium market state with the cooperating producers in a sector?
- What volume of production will each producer supply?
- What will be a common market price of a product?
- What will be a common profit?
- How will this profit be divided?

So the firms agree on a common course of action. Pepall [6] shows, that the firms will seek such a solution of equilibrium on a market that would maximize their common profit and adjust their individual production strategies to this solution. They will therefore seek optimal volumes of supply of their production based considering common profit function maximization in a form:

$$\pi(Q) = \pi\left(\sum_{i=1}^{n} Q_i\right) = \sum_{i=1}^{n} \pi_i(Q_i) = \sum_{i=1}^{n} TR_i(Q_i) - TC_i(Q_i) \tag{1}$$

where

Q_i – volume of supply of an i th oligopoly producer, $Q_i \in \mathrm{R}$,
$TC_i(Q_i)$ – real function of total costs of an i th oligopoly producer, TC_i: R→R,
$\pi_i(Q_i)$ – real profit function an i th oligopoly producer, $\pi_i(Q)$: R→R,
$TR_i(Q_i)$ – real revenue function an i th oligopoly producer, $TR_i(Q_i)$: R→R,

After expressing a price of cartel production using price-demand function P(Q) in n variables we transform a common profit function maximization problem (1) of cartel subjects to a problem in a form:

$$\pi(Q_1, Q_2, \ldots, Q_n) = \sum_{i=1}^{n} TR_i(Q_i) - TC_i(Q_i) = \sum_{i=1}^{n} \left(P\left(\sum_{i=1}^{n} Q_i\right) Q_i - TC_i(Q_i) \right) \rightarrow \max$$

where

$P(Q)$ – real price-demand function, $P(Q)$: R→R

and after a modification we get:

$$\pi(Q_1, Q_2, \ldots, Q_n) = \sum_{i=1}^{n} \left(P(Q_1, Q_2, \ldots, Q_n) Q_i - TC_i(Q_i)\right) \rightarrow \max \tag{2}$$

Optimization problem (2) represents an unconstrained common profit function maximization problem in n variables and is realized in a stationary point of a concave function of a common cartel profit satisfying necessary optimality conditions in a form:

$$\frac{\partial \pi(Q_1, Q_2, \dots, Q_n)}{\partial Q_i} = \frac{\partial \left(\sum_{i=1}^{n} \left(P(Q_1, Q_2, \dots, Q_n) Q_i - TC_i(Q_i)\right)\right)}{\partial Q_i} = 0 \qquad i = 1, 2, \dots, n \tag{3}$$

A degree of difficulty of a solution of necessary optimality conditions defined by a set of nonlinear equations (3) is, naturally, determined by a degree of difficulty of a price-demand function of a sector and cost functions of particular producers – cartel subjects.

We can show that a market price and a volume of a total supply of a sector set based on the conditions (3) are socially ineffective because they let the firms in a cartel reach higher profits while at the same time have a lower volume of supply and higher market price comparing to other forms of an oligopoly market structure. Regarding this statement let us study cartel profit maximization problem (2) as a mathematical programming problem in a following form:

$$\pi(Q_1, Q_2, \dots, Q_n) = \sum_{i=1}^{n} TR_i(Q_i) - TC_i(Q_i) = \sum_{i=1}^{n} \left(P(Q_1, Q_2, \dots, Q_n) Q_i - TC_i(Q_i)\right) \to \max$$

subject to (4)

$$Q_i \geq 0 \qquad i = 1, \dots, n$$

Let us modify optimization problem (4) to a standard form, i.e. to a problem with an objective function like this:

$$-\pi(Q_1, Q_2, \dots, Q_n) = \sum_{i=1}^{n} \left(-P(Q_1, Q_2, \dots, Q_n) Q_i + TC_i(Q_i)\right) \to \min \tag{5}$$

subject to

$$Q_i \geq 0 \qquad i = 1, \dots, n \tag{6}$$

In [8], we can see that for problems (4), (5) we can formulate generalized Lagrange function in a following form:

$$L(Q_1, Q_2, \dots, Q_n) = \sum_{i=1}^{n} -\left(P(Q_1, Q_2, \dots, Q_n) Q_i + TC_i(Q_i)\right)$$

or

$$L(Q_1, Q_2, \dots, Q_n) = -\sum_{i=1}^{n} P(Q_1, Q_2, \dots, Q_n) Q_i + \sum_{i=1}^{n} TC_i(Q_i) \tag{7}$$

Minoux [5] shows, that Kuhn-Tucker optimality conditions for a Lagrange function (7) of an optimization problem (5), (6) are formulated as follows:

$$\begin{aligned} &\frac{\partial L(Q_1, Q_2, \dots, Q_n)}{\partial Q_i} \geq 0 && i = 1, 2, \dots, n \\ &Q_i \frac{\partial L(Q_1, Q_2, \dots, Q_n)}{\partial Q_i} = 0 && i = 1, 2, \dots, n \\ &Q_i \geq 0 && i = 1, 2, \dots, n \end{aligned} \tag{8}$$

After substitution of a Lagrange function (7) to optimality conditions (8) and after another modification we get the optimality conditions in a form:

$$\begin{aligned} &\frac{-\partial\left(P(Q_i, Q_2, \dots, Q_n)\right)}{\partial Q_i} Q_i - P(Q_i, Q_2, \dots, Q_n) + MC_i(Q_i) \geq 0 && i = 1, 2, \dots, n \\ &Q_i \left(\frac{-\partial\left(P(Q_i, Q_2, \dots, Q_n)\right)}{\partial Q_i} Q_i - P(Q_i, Q_2, \dots, Q_n) + MC_i(Q_i) \right) = 0 && i = 1, 2, \dots, n \\ &Q_i \geq 0 && i = 1, 2, \dots, n \end{aligned} \tag{9}$$

If in the optimality conditions (9) we use own price elasticity of a demand for oligopoly product based on a relation

$$e(Q_i) = \frac{\Delta Q_i}{\Delta P} = \frac{\frac{\partial Q_i}{Q_i}}{\frac{\partial P(Q_i, Q_2, \dots, Q_n)}{P(Q_i, Q_2, \dots, Q_n)}}$$

than in the optimality conditions we can use a following substitution:

$$\begin{aligned} &\frac{-\partial (P(Q_i, Q_2, \dots, Q_n))}{\partial Q_i} Q_i - P(Q_i, Q_2, \dots, Q_n) = \\ &= -P(Q_i, Q_2, \dots, Q_n)\left(\frac{\partial (P(Q_i, Q_2, \dots, Q_n))}{\partial Q_i} \frac{Q_i}{P(Q_i, Q_2, \dots, Q_n)} + 1 \right) = \\ &= -P(Q_i, Q_2, \dots, Q_n)\left(\frac{1}{e(Q_i)} + 1 \right) = -P(Q_i, Q_2, \dots, Q_n) \frac{1 + e(Q_i)}{e(Q_i)} \qquad i = 1,2,\dots,n \end{aligned} \tag{10}$$

and we can finally formulate the optimality conditions in a final form:

$$-P(Q_i, Q_2, \dots, Q_n) \frac{1 + e(Q_i)}{e(Q_i)} + MC_i(Q_i) \geq 0 \qquad i = 1,2,\dots,n \tag{11.1}$$

$$Q_i \left(-P(Q_i, Q_2, \dots, Q_n) \frac{1 + e(Q_i)}{e(Q_i)} + MC_i(Q_i) \right) = 0 \qquad i = 1,2,\dots,n \tag{11.2}$$

$$Q_i \geq 0 \qquad i = 1,2,\dots,n \tag{11.3}$$

Carlton [1] shows, that the scheme would be simpler for a homogenous production and undifferentiated cost functions or the functions of marginal costs of particular producers. Now let's study the properties of cartel profit optimization problem having various marginal costs of cartel participants' homogenous production.

So if the particular producers in a cartel decide on an optimal production Q_i, for i=1, 2, ..., n, meaning that a total cartel supply and a total demand on the market are in balance while maximum of cartel profit function $\pi(Q_1, Q_2, \dots, Q_n)$, then such variables Q_i^* must exist for which the Kuhn-Tucker optimality conditions (24) are satisfied, therefore a vector of variables $(Q_1^*, Q_2^*, \dots, Q_n^*)$ is a solution to a set of equations and inequations (11.1), (11.2), (11.3).

Let us now describe some interesting, economically interpretable consequences of Kuhn-Tucker optimality conditions in a profit maximization problem in the conditions of equilibrium on a cartel market:

1. In the first case let's notice that validity of a condition (11.3) express that in a state of anticipated market equilibrium between aggregated demand and aggregated supply on a cartel market, an optimal value of supply of each cartel producer is either positive or zero. In other words, there may be producers on a market who don't get to supply their production and thus have a zero volume of supply $Q_i^* = 0$. As we show later, this is related with a fact that an equilibrium market price of a cartel covers the marginal costs of a potential cartel participant.
2. Assuming that a cartel participant supplies a positive volume of his production $Q_i^* > 0$, then from a validity of the condition (24.3) results a fact that he must supply such a volume of his production Q_i^* so that a following relation is valid between a cartel market price, his marginal costs and a demand elasticity corresponding to his market supply:

$$P(Q_i, Q_2, \dots, Q_n) = \frac{e(Q_i)}{1 + e(Q_i)} MC_i(Q_i) \tag{12}$$

Let's study an economical interpretation of a relation (12) more closely. We will show that if a company in a cartel supplies a positive volume of production then a market price of a cartel is higher than marginal costs of a firm assuming that a demand for a firm's production is elastic i.e. $e(Q_i^*) < -1$. For an elastic demand for a value of a multiplier $\frac{e(Q_i)+1}{e(Q_i)}$ in (12) stands

$$e(Q_i) < -1 \quad \rightarrow \quad \frac{e(Q_i)+1}{e(Q_i)} > 1 \tag{13}$$

and therefore also

$$P(Q_i, Q_2, \dots, Q_n) > MC_i(Q_i) \tag{14}$$

So we can see that if s firm in a cartel supplies a positive volume of its production, its marginal costs are always lower than an equilibrium market price of a cartel. Fendek [3] shows, that an optimal combination of a price and supply is under stated assumptions for a company in a cartel always more convenient than for a firm in an environment of a perfect competition which compared to (14) must supply such a value of production to equal its marginal costs with a product price.

3. From (12) we can conclude that a bigger share on cartel production is gained by the firms with lower marginal costs because with a cartel common market price they can set bigger volumes of their supply having lower marginal costs. This positive trend of a firm's placement in a cartel is even more significant in a case if higher level of demand elasticity for a firm's production.
4. In a context of above said notes let's discuss economical interpretation of the Kuhn-Tucker optimality condition (11.1). Two situations may occur:
   a. In a case a firm in a point of cartel optimal equilibrium supplies a positive volume of production $Q_i^* > 0$ then from a validity of the condition (11.2) results that a cartel market price is on a level of marginal costs of a firm multiplied by the multiplier from (25) and the optimality condition (24.1) is then realized as an equation and:
   $$-P(Q_i, Q_2, \ldots, Q_n)\frac{1+e(Q_i)}{e(Q_i)} + MC_i(Q_i) = 0$$
   b. In other case, i.e. if the optimality condition (11.1) is realized as a sharp inequality and
   $$-P(Q_i, Q_2, \ldots, Q_n)\frac{1+e(Q_i)}{e(Q_i)} + MC_i(Q_i) > 0$$
   then from a validity of the condition (11.2) formally results that a company in a point of cartel optimal equilibrium is not producing, i.e. volume if production is $Q_i^* = 0$. Vivies [7] shows, that this situation occurs because the firm appears to be technologically ineffective as it has high marginal costs so these after being multiplied by the multiplier from (12) are not covered by a cartel market price
   $$P(Q_i, Q_2, \ldots, Q_n) < \frac{e(Q_i)}{1+e(Q_i)} MC_i(Q_i)$$

Kuhn-Tucker optimality conditions for cartel profit maximization problem confirmed a fact that the firms in a cartel, because of the existing barriers to entry and using their agreement on joint course of action in setting an optimal combination of a price and a volume of product supply on a sector market, set the cartel product price above the level of marginal costs and therefore they reach a superior profit.

As an analysis of Kuhn-Tucker optimality conditions has clearly proven, such situation is a result of a fact that the firms acting on a sector market in a cartel structure formally declare their relations being competitive but in fact they act as a monopoly while setting a combination of a market price and a cartel supply, with only one difference of a profit being then divided by a scheme based on comparison of their marginal costs as it was shown by the description of (11.2) Kuhn-Tucker optimality conditions implications.

3 Conclusion

For current developed economies the different forms of imperfect competition are typical as a prevailing type of market structures. Among these a significant position belongs to an oligopoly which is characteristic for most sectors of national economies of developed countries. In an economic theory, a great attention is paid to the study of theoretical concepts of oligopolies behavior also in a context of applying the principles and rules of economic competition on the markets of those sectors where a strong economic subject exists.

That is because if the objective barriers to entry exist on a relevant market, the existence of economic subjects with a dominant market share is clearly connected with a high risk of a dominant position misuse. Legislation for guarantee of competition conditions and elimination of risks of anti-competition practices is created in the developed economies. In Slovakia these tasks are performed by the Antimonopoly Office of the Slovak Republic.

An oligopoly represents a market structure where a limited number of producers operate on a market of a sector. Firms in an oligopoly must respect an existence of their competitors and seek mutual agreement based on different assumptions and schemes of aligning the interests which will be determined by specific characteristics of a sector. In

this competition scheme an existence of an oligopoly in accordance with the rules of economic competition is socially effective and a presence of the competitive relations on a producers' side is naturally effective as well.

A different situation occurs when the firms in an oligopoly which formally declare their independency and the existence of competitive relations in a sector, close a secret agreement on joint process of fixing a volume of supply and price of a product in a sector, which is clearly against the principles of a competition protection. In this case we speak of a cartel. An aim of this practice is an effort to reach an extra profit for cartel subjects.

If the producers in a sector close the agreements based on cartel principles, they significantly limit a quality of a competition and it is an obligation of a state to regulate or eliminate this condition by legislation. In this paper we dealt with methodological tools of microeconomic analysis for fixing an optimal supply and price of a cartel and showed the ineffectiveness of such market structure.

References

[1] Carlton, D.W.; Perloff, J. M.: Modern Industrial Organization. Boston: Addison Wesley, 2005. ISBN 0-321-22341-1, 2005.

[2] Fendek, M.; Fendeková, N.: Models of Price Regulation of Network Industries. Ekonomický časopis: časopis pre ekonomickú teóriu a hospodársku politiku, spoločensko-ekonomické prognózovanie. Bratislava Prognostický ústav SAV, 2010. ISSN 0013-3035, 2010, roč. 58, č. 10, 2010.

[3] Fendek, M.: Natural monopoly cost-oriented price regulation. Quantitative methods in economics: multiple criteria decision making XIV. - Bratislava: IURA EDITION, 2008. ISBN 978-80-8078-217-7, 2008.

[4] Fendek, M. – Fendekova, E.: Application of the price regulation models in the network industries market in Slovakia. 23rd European Conference on Operational Research. Bonn, Deutschland, July 5-8, 2009.

[5] Minoux, M.: Mathematical Programming (Theory and Applications). John Wiley & Sons, New York 1986.

[6] Pepall, L. – Richards, D. - Norman , D.: Industrial Organization: Contemporary Theory and Empirical Applications.New York: Wiley-Blackwell, 2008.

[7] Vivies, X.: Oligopoly Pricing. Old Ideas and New Tools. Cambridge, New York: The MIT Presss, 2007.

[8] Williams, H.P.: Model Building in Mathematical Programming. John Wiley and Sons, London 2003.

[9] Wiscusi, W. K. – Vernon, J. M. – Harrington, J. E.: Economics of Regulation and Antitrust. Cambridge: The MIT Press, 2004

Modeling of competition in revenue management

Petr Fiala[1]

Abstract. Revenue management (RM) is the art and science of predicting consumer behavior and optimizing price and product availability to maximize revenue. RM models, despite their success and popularity, still remain somewhat simplistic. The most critical flaw is that current RM models are designed under the assumption that demands are independent random variables. The models are readily extended to competitive settings. The paper is devoted to modeling of competition in revenue management. The revenue management problems under competition can be formulated as games. Results from game theory allow to study the existence and uniqueness of equilibrium policies in revenue management games. An approximation algorithm can be used for solving the revenue management problems under competition.

Keywords: network revenue management, competition, game theory.

JEL Classification: C44
AMS Classification: 90B50

1 Introduction

Revenue management (RM) is the art and science of predicting consumer behavior and optimizing price and product availability to maximize revenue (see [6], [8]). RM models, despite their success and popularity, still remain somewhat simplistic. The most critical flaw is that current RM models are designed under the assumption that demands are independent random variables. Many important components such as pricing, capacity management, overbooking, network revenue management, and choice modeling have been extensively studied, competition have not received enough attention. The models are readily extended to competitive settings.

Network revenue management models attempt to maximize revenue when customers buy bundles of multiple resources. The dependence among the resources in such cases is created by customer demand. This class of problems arises for example in airlines, the problem is managing capacities of a set of connecting flights across a network. The basic model of the network revenue management problem is formulated as a stochastic dynamic programming problem whose exact solution is computationally intractable. There are several approximation methods for the problem. The Deterministic Linear Programming (DLP) method is a popular in practice. The DLP method is based on an assumption that demand is deterministic and static.

Depending on the chosen decision variables, competition in revenue management can be categorized as either price-based or quantity-based. Competition in models can be classified according to some characteristics, as static and dynamic, horizontal and vertical, and others. The revenue management problems under competition can be formulated as games. Results from game theory allow to study the existence and uniqueness of equilibrium policies in revenue management games. A comparison between the centralized system and the decentralized system is a traditional topic for studying game-theoretic models. In the paper [4], authors consider horizontal competition over a single-leg flight and vertical competition over a series of connecting fights, assuming low-fare passengers arrive earlier than high-fare passengers. They compare the centralized and decentralized (competitive) settings. In the paper [3], authors have studied airline capacity allocation under competition. Using the concept of demand overflow, they have proposed game-theoretic models based on well-known approximate models for network capacity allocation in the monopolistic setting. They have investigated both the existence and uniqueness of a Nash equilibrium for different game-theoretic models.

The paper is devoted to modeling of competition in network revenue management. The paper analyzes linear approximations of the joint problem of competition and capacity control in network revenue management. The model combines the DLP model with a game model of competition. The rest of the paper is organized as follows. In section 2, network revenue management models and the DLP method are summarized. In Section 3, the DLP network model is extended to include competition. In Section 4, the competition model is analyzed as a generalized Nash game and a Nash game. In the last section, some concluding remarks to possible extensions and to further research are made.

[1] University of Economics, Department of Econometrics, W. Churchill Sq. 4, 13067 Prague 3, pfiala@vse.cz.

2 Network revenue management

The quantity-based revenue management of multiple resources is referred as network revenue management. This class of problems arises for example in airline, hotel, and railway management. In the airline case, the problem is managing capacities of a set of connecting flights across a network, so called a hub-and-spoke network. In the hotel case, the problem is managing room capacity on consecutive days when customers stay multiple nights.

Network revenue management models attempt to maximize some reward function when customers buy bundles of multiple resources. The interdependence of resources, commonly referred to as network effects, creates difficulty in solving the problem. The classical technique of approaching this problem has been to use a deterministic LP solution to derive policies for the network capacity problem. Initial success with this method has triggered considerable research in possible reformulations and extensions, and this method has become widely used in many industrial applications. A significant limitation of the applicability of these classical models is the assumption of independent demand. In response to this, interest has arisen in recent years to incorporate customer choice and competition into these models, further increasing their complexity. This development drives current efforts to design powerful and practical heuristics that still can manage problems of practical scope.

The basic model of the network revenue management problem can be formulated as follows (see [6], [8]): The network has m resources which can be used to provide n products. We define the incidence matrix $\mathbf{A} = [a_{hk}]$, $h = 1, 2, \ldots, m$, $k = 1, 2, \ldots, p$, where

$a_{hk} = 1$, if resource h is used by product k, and

$a_{hk} = 0$, otherwise.

The k-th column of $\mathbf{A}$, denoted $\mathbf{a}_k$, is the incidence vector for product k. The notation $h \in \mathbf{a}_k$ indicates that resource h is used by product k.

The state of the network is described by a vector $\mathbf{c} = (c_1, c_2, \ldots, c_m)$ of resource capacities. If product k is sold, the state of the network changes to $\mathbf{c} - \mathbf{a}_k$.

Time is discrete, there are T periods and the index t represents the current time, $t = 1, 2, \ldots, T$. Assuming within each time period t at most one request for a product can arrive.

Demand in time period t is modeled as the realization of a single random vector $\mathbf{r}(t) = (r_1(t), r_2(t), \ldots, r_p(t))$. If $r_k(t) = r_k > 0$, this indicates a request for product k occurred and that its associated revenue is r_k. If $r_k(t) = 0$, this indicates no request for product k occurred. A realization $\mathbf{r}(t) = \mathbf{0}$ (all components equal to zero) indicates that no request from any product occurred at time t. The assumption that at most one arrival occurs in each time period means that at most one component of $\mathbf{r}(t)$ can be positive. The sequence $\mathbf{r}(t)$, $t = 1, 2, \ldots, T$, is assumed to be independent with known joint distributions in each time period t. When revenues associated with product k are fixed, we will denote these by r_k and the revenue vector $\mathbf{r} = (r_1, r_2, \ldots, r_p)$.

Given the current time t, the current remaining capacity $\mathbf{c}$ and the current request $\mathbf{r}(t)$, the decision is to accept or not to accept the current request. We define the decision vector $\mathbf{u}(t) = (u_1(t), u_2(t), \ldots, u_p(t))$ where

$u_k(t) = 1$, if a request for product k in time period t is accepted, and

$u_k(t) = 0$, otherwise.

The components of the decision vector $\mathbf{u}(t)$ are functions of the remaining capacity components of vector $\mathbf{c}$ and the components of the revenue vector $\mathbf{r}$, $\mathbf{u}(t) = \mathbf{u}(t, \mathbf{c}, \mathbf{r})$. The decision vector $\mathbf{u}(t)$ is restricted to the set

$$U(\mathbf{c}) = \{\mathbf{u} \in \{0,1\}^n,\ \mathbf{Au} \leq \mathbf{c}\}.$$

The maximum expected revenue, given remaining capacity $\mathbf{c}$ in time period t, is denoted by $V_t(\mathbf{c})$. Then $V_t(\mathbf{c})$ must satisfy the Bellman equation

$$V_t(\mathbf{c}) = E\left[\max_{u \in U(\mathbf{c})} \{\mathbf{r}(t)^T \mathbf{u}(t, \mathbf{c}, \mathbf{r}) + V_{t+1}(\mathbf{c} - \mathbf{Au})\}\right] \tag{1}$$

with the boundary condition

$$V_{T+1}(\mathbf{c}) = 0, \forall \mathbf{c}.$$

A decision $\mathbf{u}^*$ is optimal if and only if it satisfies:

$u_j(t, \mathbf{c}, r_j) = 1$, if $r_j \geq V_{t+1}(\mathbf{c}) - V_{t+1}(\mathbf{c} - \mathbf{a}_j)$, $\mathbf{a}_j \leq \mathbf{x}$,

$u_j(t, \mathbf{c}, r_j) = 0$, otherwise.

This reflects the intuitive notion that revenue r_k for product k is accepted only when it exceeds the opportunity cost of the reduction in resource capacities required to satisfy the request. The equation (1) cannot be solved exactly for most networks of realistic size. Solutions are based on approximations of various types. There are two important criteria when judging network approximation methods: accuracy and speed.

The first approach is to use a simplified network model, for example posing the problem as a static mathematical program. We introduced Deterministic Linear Programming (DLP) method (see [8]).

The DLP method uses the approximation

$$\max_{\mathbf{x}} \mathbf{r}^T\mathbf{x}$$

subject to (2)

$$\mathbf{Ax} \leq \mathbf{c},$$
$$\mathbf{x} \leq \mathbf{D},$$
$$\mathbf{x} \geq \mathbf{0},$$

where $\mathbf{D} = (D_1, D_2, \ldots, D_p)$ is the vector of demand over the periods $t, t+1, \ldots, T$, for product k, $k = 1, 2, \ldots, p$, and $\mathbf{r} = (r_1, r_2, \ldots, r_p)$ is the vector of revenues associated with the p products. The decision vector $\mathbf{x} = (x_1, x_2, \ldots, x_p)$ represent partitioned allocation of capacity for each of the p products. The approximation effectively treats demand as if it were deterministic and equal to its mean $E[D]$.

The optimal dual variables, $\boldsymbol{\pi}^{\mathrm{LP}}$, associated with the constraints $\mathbf{Ax} \leq \mathbf{c}$, are used as bid prices. The DLP was among the first models analyzed for network RM. The main advantage of the DLP model is that it is computationally very efficient to solve. Due to its simplicity and speed, it is a popular in practice. The weakness of the DLP approximation is that it considers only the mean demand and ignores all other distributional information. The performance of the DLP method depends on the type of network, the order in which fare products arrive and the frequency of re-optimization.

3 Modeling of competition

Modeling of competition among n firms, indexed $i = 1, 2, \ldots, n$, is based on optimization models of DLP type for individual firms. The next step is searching for equilibrium for competing firms.

We make the following standard assumptions (see [3]):

- The prices of all products are fixed for all firms.
- The demand for one product is independent of that for another product.
- The demand for one product from one firm is correlated to the demand for the same product from other firms.
- Each customer is interested only in one particular product.
- Each customer makes a booking request from his preferred firm and with a certain probability, makes another booking request of the same product from another firm if his first booking request is rejected. If his second booking request is also rejected, then he becomes a lost customer to all firms for this time.

Firms sell p classes of products, indexed $k = 1, 2, \ldots, p$, combined from m resources, indexed $h = 1, 2, \ldots, m$. Let $\mathbf{r}^i$ be unit price vector for firm i. Let $\mathbf{c}^i$ be remaining capacity vector for firm i. Let A^i be the resource-product incidence matrix for firm i. Assume that primary demand for firm i is $\mathbf{D}^i$. A rejected customer from firm i makes another booking request for the same product from other firms. Suppose d^{ji} denotes the overflow rate of particular product from firm j to firm i. That is, if a customer, who prefers firm j, is rejected for a booking request for product by firm j, then he would make a booking request of product from firm i with a probability d^{ji}. The total potential demand for firm i is made up from its own primary demand and the overflow demand from other firms, $\mathbf{D}^i + \sum_{j \neq i} d^{ji} [\mathbf{D}^j - \mathbf{x}^j]^+$.

Assume that partitioned booking limits for all other firms other than i are given, firm i aims to determine its optimal partitioned booking limits $\mathbf{x}^i$ by solving the following deterministic linear program (DLP):

$$\max_{\mathbf{x}^i} (\mathbf{r}^i)^T\mathbf{x}^i$$

subject to (3)

$$\mathbf{A}^i\mathbf{x}^i \leq \mathbf{c}^i,$$
$$\mathbf{x}^i \leq \mathbf{D}^i + \sum_{j \neq i} d^{ji} [\mathbf{D}^j - \mathbf{x}^j]^+,$$
$$\mathbf{x}^i \geq \mathbf{0}.$$

Each firm satisfies its primary demand and then accepts the overflow demand that cannot be satisfied by rival firms. The objective is that firm i maximizes its total revenue. The first constraints states that the capacity on each resource must not be violated. The second constraint specifies that the allocation to all firms for each product must not exceed the demand for this product. The last constraint shows that the booking limits are nonnegative.

There is possible reformulate the problem (3) into an equivalent nonlinear and non-smooth problem, whose feasible set depends only on the partitioned booking limit $\mathbf{x}^i$ of firm i.

Let $\mathbf{s}^i > \mathbf{r}^i$ be a constant vector for any i.

$$\max_{\mathbf{x}^i} (\mathbf{r}^i)^T \mathbf{x}^i + (\mathbf{s}^i)^T \min\left(0, \mathbf{D}^i + \sum_{j \neq i} d^{ji} [\mathbf{D}^j - \mathbf{x}^j]^+ - \mathbf{x}^i\right)$$

subject to (4)

$$\mathbf{A}^i \mathbf{x}^i \leq \mathbf{c}^i,$$
$$\mathbf{x}^i \geq \mathbf{0}.$$

The vector $\mathbf{x}^i$ is an optimal solution to the problem (3) if and only if $\mathbf{x}^i$ is an optimal solution to the problem (4). The feasible set of the problem (4) is simpler that of the problem (3).

4 Game models

In this section, we introduce generalized Nash games and generalized Nash equilibrium points (see [5]). The relationship between competition models and game models is shown. Next the existence and uniqueness of generalized Nash equilibrium points are studied. An outline of algorithm for solving competition models is given.

The **generalized Nash game** is a non-cooperative game in which each player's admissible strategy set depends on the other players' strategies. Assume that there are n players and each player i, $i = 1, 2, \ldots, n$, controls variables $\mathbf{x}^i$. In fact $\mathbf{x}^i$ is a strategy of the player i.

Let denote by $\mathbf{x}$ the following vector

$$\mathbf{x} = (\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n).$$

And let $N = n_1 + n_2 + \ldots + n_n$. Thus $\mathbf{x} \in R^N$.

Denote by $\mathbf{x}^{-i}$ the vector formed of all players' decision variables except the one of the player i. So we can write

$$\mathbf{x} = (\mathbf{x}^i, \mathbf{x}^{-i}).$$

The strategy of the player i belongs to a strategy set

$$\mathbf{x}^i \in X^i(\mathbf{x}^{-i})$$

which depends on the decision variables of the other players.

Let $f^i(\mathbf{x}^i, \mathbf{x}^{-i})$ be the payoff function for player i when the joint strategy is $\mathbf{x}$. Aim of the player i, given the strategy $\mathbf{x}_0^{-i}$, is to choose a strategy $\mathbf{x}_0^i$ that solves the following optimization problem

$$\max_{\mathbf{x}^i} f^i(\mathbf{x}^i, \mathbf{x}_0^{-i})$$

subject to (5)

$$\mathbf{x}^i \in X^i(\mathbf{x}^{-i}).$$

For any given strategy vector $\mathbf{x}^{-i}$ of the rival players the solution set of the problem (5) is denoted by $S^i(\mathbf{x}^{-i})$.

Thus a vector $\mathbf{x}_0$ is a **generalized Nash equilibrium** if for any i,

$$\mathbf{x}_0^i \in S^i(\mathbf{x}_0^{-i}).$$

Whenever the strategy set of each player does not depend on the choice of the rival players, that is, for any i,

$$X^i(\mathbf{x}^{-i}) = X^i$$

then the non-cooperative game reduces to find $\mathbf{x}_0 \in \prod_i X^i$ that for any i,

$$f^i\left(\mathbf{x}_0^i, \mathbf{x}_0^{-i}\right) = \max_{\mathbf{x}^i} f^i\left(\mathbf{x}^i, \mathbf{x}_0^{-i}\right)$$

subject to (6)

$$\mathbf{x}_0^i \in X^i,$$

that is a **Nash game**.

If there is no joint constraint in the game, then the generalized Nash game and a generalized Nash equilibrium reduce to a traditional Nash game and a Nash equilibrium respectively. The key difference between generalized Nash games and traditional Nash games is that the strategy space for a player may depend on other players' strategies in the former, but not in the latter, although the payoff functions in both types of games are allowed to be functions of other players' strategies.

The network revenue management DLP problems (3) and (4) are models of competition. The feasible set (strategy set) of the problem (3) involves the strategy variables $\mathbf{x} = (\mathbf{x}^i, \mathbf{x}^{-i})$, while the feasible set (strategy set) of the problem (4) only involves the strategy variables $\mathbf{x}^i$. It is easy to observe that the game based on DLP problem (3) for all firms results in a generalized Nash game, while the game based on DLP problem (4) results in a traditional Nash game with non-smooth and nonlinear payoff functions. This non-smooth property may pose difficulties for proposing computational methods for solving games. DLP problems (3) and (4) are equivalent. Therefore, in the context of game theory, a generalized Nash game is converted into a traditional Nash game.

The existence of a Nash equilibrium (or a generalized Nash equilibrium) for Nash games (or generalized Nash games) is an important topic in game theory. Without equilibrium in a game, players do not know what strategy they should take. The uniqueness of the Nash equilibrium is another important topic in game theory. If there is a unique equilibrium, players can choose their strategies without vagueness. The obvious problem with multiple equilibria is that the players may not know which equilibrium will prevail. The main results for the revenue management games are given in Lemma 1, Theorem 1, and Remark 1.

Lemma 1. *Vector* $\mathbf{x}_0$ *is a generalized Nash equilibrium for the generalized Nash game defined by DLP problem* (3) *if and only if* $\mathbf{x}_0$ *is a Nash equilibrium for the Nash game defined by DLP problem* (4).

Proof. The result follows from equivalence between problems (3) and (4) and the definitions of the generalized Nash equilibrium and Nash equilibrium.

Theorem 1. *There exists a generalized Nash equilibrium for the game based on DLP problem* (3).

Proof. The result follows from Lemma 1 and Theorem 1 of [7], which states that a Nash equilibrium exists for a Nash game if the payoff function for each player is concave with respect to their own strategy and continuous with respect to the strategies of all players and the strategy set for each player is convex and compact.

Remark 1. The uniqueness of a generalized Nash equilibrium for the game based on DLP problem (3) is not guaranteed. There is a unique equilibrium result for traditional Nash game (see [1]). In order to obtain the uniqueness of a Nash equilibrium, the payoff function must be twice continuously differentiable with respect to all strategy variables. Since the payoff functions in Nash game based on DLP problem (4) are non-smooth, the uniqueness of the Nash equilibrium is not guaranteed. Result holds with regard to Lemma 1.

It is well known that traditional Nash games are equivalent to variational problems when the payoff function for each player is continuously differentiable and concave to its own strategies. (see [1]). Generalized Nash games are equivalent to quasi-variational inequality problems. Algorithms for solving quasi-variational problems are not common in the literature. A sequential penalty method for a general quasi-variational inequality problem is proposed in [5]. Another simple approximation algorithm is proposed to use for solving generalized Nash games, where it is not applied a penalty approach but DLP problems are solved in each iteration.

Approximation algorithm

Step 1 (Initialization)
Choose the stopping rule parameter ϵ. Let $s = 1$. Choose a starting point

$$\mathbf{x}(s) = (\mathbf{x}^1(s), \mathbf{x}^2(s), \ldots, \mathbf{x}^n(s)).$$

Step 2 (Searching)
For each player i at iteration s+1, finding $\mathbf{x}^i(s+1)$ by solving an optimization DLP problem (3) assuming

$$\mathbf{x}^{-i}(s+1) = (\mathbf{x}^1(s+1), \ldots, \mathbf{x}^{i-1}(s+1), \mathbf{x}^{i+1}(s), \ldots, \mathbf{x}^n(s))$$

is given.

Step 3 (Stopping rule)
If it holds

$$\|\mathbf{x}(s+1) - \mathbf{x}(s)\| \leq \epsilon,$$

then the algorithm terminates and $\mathbf{x}(s)$ is an approximate generalized Nash equilibrium. Otherwise, set $s := s+1$ and go to Step 2.

5 Conclusions

Revenue management (RM) models, despite their success and popularity, still remain somewhat simplistic. Many important components such as pricing, capacity management, overbooking, network revenue management, and choice modeling have been extensively studied, competition have not received enough attention. The models are readily extended to competitive settings.

In the paper, network revenue management models are completed with game models of competition. The approach seems to be useful and promising for next research. There are some possible extensions of the approach and some areas for further research. The approach can be adapted to other network approximations as well (such as RLP, PNLP, and others) (see [2]). The same formulation can be applied to a variety of network bid-price methods. While it is easy to prove the existence of a Nash or generalized Nash equilibrium, it appears difficult to ensure the uniqueness in general. Some special cases should be analyzed. Convergence of the approximation algorithm for generalized Nash games is a challenge for a future investigation.

Acknowledgements

The research project was supported by Grant No. 402/10/0197 „Revenue management – models and analyses“ from the Grant Agency of the Czech Republic and by Grant No. IGA F4/16 /2011, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Facchinei, F., and Pang, J.: *Finite-Dimensional Variational Inequalities and Complementarity Problems I and II*. Springer, New York, 2003.
[2] Fiala, P.: Approximations for solving network revenue management problems. In: *Mathematical Methods in Economics 2010*. University of South Bohemia, České Budějovice, 2010, 132–137.
[3] Jiang, H., and Pang, Z.: Network capacity management under competition. *Computational Optimization and Applications* **50** (2011), 287–326.
[4] Netessine, S., and Shumsky, R.: Revenue management games: horizontal and vertical competition. *Management Science* **51** (2005), 813–831.
[5] Pang, J., and Fukushima, M.: Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games. *Computational Management Science* **1**: (2005), 21–56.
[6] Phillips, R.: *Pricing and Revenue Optimization*. Stanford Business Books, Stanford, California, 2005.
[7] Rosen, J.: Existence and uniqueness of the equilibrium points for concave n-person games. *Econometrica* **33** (1965), 520–534.
[8] Talluri, K. T., and van Ryzin, G. J.: *The Theory and Practice of Revenue Management*. Kluwer Academic Publishers, Boston, MA, 2004.

Optimizing Permutation Methods for the Ordinal Ranking Problem

Jan Fiedor[1], Jiří Mazurek[2]

Abstract. In many real-world decision-making situations and preferential votings objects or alternatives are ranked by experts from the best to the worst with the goal to achieve a group consensus. This setting is called ordinal (consensus) ranking problem (OCRP). Many methods were proposed for the solution of OCRP with an important class of the so called permutation methods such as the consensus ranking model (CRM) or hybrid distance-based ideal-seeking consensus ranking model (DCM), which seek the consensus in the space of all permutations of the order n. The main limitation of these methods is that the number of permutations of order n grows as n!, making solving this problem virtually impossible for n larger than 10. The aim of the article is to propose a new technique which can reduce the number of permutations the methods' need to process, lowering their computational complexity and enabling them to deal with permutations of higher orders. The technique is based on the 'distance theorem' derived in the theoretical part of the paper. This theorem postulates the upper limit on the distance between one of the initial preferences (called median) and the consensus. In the empirical part of the paper, Monte Carlo simulations are carried out to examine the technique's behaviour, applicability and efficiency.

Keywords: consensus, group decision making, Monte Carlo simulations, ordinal preference rankings, permutations.

JEL classification: D71
AMS classification: 65K99, 91B06, 91B08, 91B14

1 Introduction

In business, politics, entertainment or sport we encounter situations involving preferential votings of objects or alternatives. Often, such preferences have structure of *ordinal ranking*, where objects are assigned their rank order from the 1st to the n^{th} place. This is common especially when a set of experts compile various TOP 10, 25, 100, etc., or when some alternatives in a decision-making situation have to be sorted from the best to the worst. In this setting, finding a group consensus is one of the most important and also difficult tasks, which dates back to the late 18th century and works by Borda [2] and Condorcet [3]. As shown by Arrow in [1] it is impossible to find a unique solution (consensus) to all preference ranking problems while meeting a set of some reasonable criteria (e.g. non-dictatorship).

Problems, which have a solution, can be solved through many different methods proposed over the last centuries (see [2, 3]) as well as recently ([5, 10, 13]). Probably the most familiar methods are Condorcet's simple majority rule and Borda-Kendall's method of marks. Modern methods treat preferences as vectors, permutations or matrices, and use distance functions defined on vector or matrix spaces ([4, 11]), but generally with different results ([6, 13]). Important class of these methods constitute so called *permutation methods* such as the *consensus ranking model* (CRM) of Cook and Kress [5] or the *hybrid distance-based ideal-seeking consensus ranking model* (DCM) by Tavana et al. [13]. These methods seek the consensus in the space of all permutations of order n. The consensus is defined as a preference (permutation) with the minimal sum of distances to preferences of all DMs. The limitation of these methods is that the number of permutations of order n grows as $n!$, making solving this problem virtually impossible for n larger than 10.

The aim of the article is to propose a new technique for permutation methods which can significantly reduce the number of permutations in consideration, thus reducing their computational complexity. The technique is based on the 'distance theorem' derived in the theoretical part of the paper. In the empirical part, Monte Carlo simulations of the technique are carried out to examine its behaviour, applicability and efficiency.

[1] Brno University of Technology, Faculty of Information Technology, Department of Intelligent Systems, DITS FIT BUT, Božetěchova 2, 612 66 Brno, e-mail: ifiedor@fit.vutbr.cz.
[2] Silesian University in Opava, School of Business Administration in Karviná, Department of Mathematical Methods in Economics, Univerzitní náměstí 1934, 733 40 Karviná, e-mail: mazurek@opf.slu.cz.

The article is organized as follows. In Section 2, basic concepts are introduced. In Section 3, the distance theorem is derived and the technique based on this theorem is presented in Section 4. Numerical examples are provided in Section 5 and computer simulations are performed in Section 6. Conclusions close the article.

2 Concepts and notation

Let a set of n alternatives be ranked by a set of k decision makers (DMs). Each DM provides the ranking of all alternatives from the 1st to the n^{th} position. Hence, the number of decision makers is equal to the number of preference rankings, i.e., each DM provides only one ranking. All rankings have the same weight. Rankings of alternatives, such as (A, C, D, B), can be converted into (numeric) permutations by a simple correspondence: $A \equiv 1$, $B \equiv 2$, etc., leading to the permutation (1, 3, 4, 2). In this paper, permutations are denoted $\pi_i, i \in \{1,..,k\}$ and the set of all permutations of order n as S_n.

Because permutations of different DMs might be equal, the set of DMs' permutations is in fact a *multiset* (bag) (A, m) with the *underlying set of elements* A and the *multiplicity function* $m : A \to N$.

Furthermore, we assume that there exist a *metric function* d on the set A. Generally, a metric on a set A is a function $d : A \times A \to \mathbf{R}$, which satisfies the following conditions:

1. $d(x, y) \geq 0$
2. $d(x, y) = 0$ if and only if $x = y$
3. $d(x, y) = d(y, x)$
4. $d(x, z) \leq d(x, y) + d(y, z)$, for $x, z, y \in A$.

Some appropriate metrics for permutations (derived from metric functions on vector spaces) used in this paper are listed below.

- *Kendall's tau distance* ([7, 8]): the number of transposition of adjacent pairs of digits necessary to turn one permutation into other. Also known as *bubble-sort distance*.
- L_p *metric*: distance of two permutations π_i and π_j is given as:

$$d(\pi_i, \pi_j) = \sqrt[p]{\sum_{k=1}^{n} \left| \pi_{ik} - \pi_{jk} \right|^p} \tag{1}$$

  For $p = 1$ we obtain the *Manhattan metric* and for $p = 2$ the *Euclidean metric*.
- As permutations can be converted into permutation matrices, also metric functions on matrix spaces analogous to (1) are available. Such l_1 metric is used in DCM and CRM methods.

Definition 1. Let (A, m) be a multiset of permutations given by all DMs, and d be a metric function. Then:

- *Distance between two permutations* $\pi_i \in A$ *and* $\pi_j \in A$ *is denoted* $d(\pi_i, \pi_j)$.
- *Median is a permutation* $\pi_m \in A$ *with the lowest overall distance* $M = \sum_{i=1}^{k} d(\pi_m, \pi_i)$ *to all permutations from A. Generally, there can be more than one median.*
- *Consensus is a permutation* $\pi_r \in S_n$ *with the lowest overall distance* $\sum_{i=1}^{k} d(\pi_r, \pi_i)$ *to all permutations from A. A consensus can belong to A as well, and there can be more than one consensus too.*

n / k	0	1	2	3	4	5	6	7	8	9	10
1	1										
2	1	1									
3	1	2	2	1							
4	1	3	5	6	5	3	1				
5	1	4	9	15	20	22	20	15	9	4	1
10	1	9	44	155	440	1068	2298	4489	8095	13640	21670
…	…	…	…	…	…	…	…	…	…	…	…

n / k (continued)	11	12	13	14	15	16	17	18	19	20	…
10	32683	47043	64889	86054	110010	135853	162337	187959	2011089	230131	…
…	…	…	…	…	…	…	…	…	…	…	…

Table 1 Distribution of permutations of order n (vertical axis) with k inversions (horizontal axis) (source: [9]).

As settings with more than one median might cause problems, we also define the *well-defined problems* and *ill-defined problems*:

Definition 2. *Well-defined problem with ordinal preference rankings is a problem where there is precisely one median in a bag A, otherwise it is an ill-defined problem.*

For better understanding of the presented concepts and the *distance theorem* in the next section we provide several simple examples in Section 5. We already know that there are *n*! permutations of order *n*, and the maximum distance is $\frac{n(n-1)}{2}$. Another important feature of permutations is their distribution with regard to the number of transpositions of adjacent pairs (inversions) to a given permutation, see Table 1. The distribution is symmetric and also asymptotically normal. Generating functions can be found in [9] and [12].

3 The distance theorem

Intuitively, a consensus seems to be 'close' to the median, as median is a permutation 'in the middle' of DMs' preferences. If this is true, it suffices to search through permutations close to the median (with only a few inversions), while permutations more distant can be safely ruled out. This, in turn, might drastically narrow the space S_n we search through. In this section, we provide a theoretical upper boundary on the distance between a median and a consensus using an arbitrary metric function *d*.

Theorem 1. Let π_m be a median permutation and let π_r be a consensus permutation, let *k* be the number of permutations and let $M = \sum_{i=1}^{k} d(\pi_m, \pi_i)$. Then $d(\pi_m, \pi_r) \leq \frac{2M}{k}$.

Proof: According to Definition 1, the median is the permutation which satisfies $M = \sum_{i=1}^{k} d(\pi_m, \pi_i) \leq \sum_{i=1}^{k} d(\pi_j, \pi_i)\ \forall j \in \{1,..,k\}$ and for the consensus $\sum_{i=1}^{k} d(\pi_r, \pi_i) \leq M$ holds. Using the triangle inequality in the definition of a metric we get:

$$d(\pi_m, \pi_r) \leq d(\pi_m, \pi_i) + d(\pi_i, \pi_r)\ \forall i \in \{1,..,k\},$$

and after summation over all *i* we obtain:

$$k \cdot d(\pi_m, \pi_r) \leq \sum_{i=1}^{k} d(\pi_m, \pi_i) + \sum_{i=1}^{k} d(\pi_i, \pi_r) \leq 2M,$$

after dividing *k* we finally get to the result: $d(\pi_m, \pi_r) \leq \frac{2M}{k}$. Q.E.D.

Definition 2. *The neighbourhood of the median (N) is a space of* $\pi \in N \subseteq S_n$ *such that* $d(\pi_m, \pi) \leq \frac{2M}{k} = r$, *where r is the radius of N.*

Corollary: $\pi_r \in N$.

The theorem narrows the searched space from S_n to *N*. When radius *r* is high, the cardinality of subspace *N* is not significantly lower than that of S_n. However, when *r* is low, then the subspace *N* might be so small that the consensus can be easily found even for a very large *n*. To illustrate the point, consider a situation where $n = 10$, $k = 6$ and $M = 30$. In permutation methods, such as CRM or DCM, a consensus is searched through S_n with *n*! = 3,628,800 permutations. However, according to the distance theorem, it suffices to search only through permutations closer than or equal to 10 inversions to the median, which is (see Table 1) only about 52,000 cases. Thus, there is a reduction of more than 3,5 million permutations (the factor of 70)!

4 Reduction technique

To allow one to solve problems involving permutations of higher orders using the permutation methods, we propose a technique for reducing the amount of permutations which the methods need to search in order to lower their overall computational complexity. The reduction technique exploits the distance theorem presented in the previous section to narrow the searched space from S_n to *N*. It alters the computation performed by a permutation method so that it proceeds in the following steps:

1. Preference rankings of *n* alternatives by *k* DMs are turned into permutations.
2. Median is found as the permutation(s) with the lowest overall distance to all the preference rankings.
3. If there is more than one median, the problem is ill-defined and the computation stops.
4. The radius *r* is computed to determine the space (*N*) which needs to be searched.
5. The efficiency of the reduction is tested and the computation stops if the subspace *N* is still too large to find the consensus in a reasonable time.
6. A consensus is found in the neighbourhood (*N*) of median using an arbitrary permutation method.

The metric used in step 2 is arbitrary. The step 3 excludes ill-defined problems explicitly, but cases with two or more medians can be, in general, handled too – by proceeding into step 6 for each median separately. Note that performing the reduction is quite cheap as its computation complexity is n^2k^2, which imposes a negligible overhead compared to the *n*! iterations a permutation method normally needs to perform, so even if the reduction is not significant, there is no reason not to try it.

5 Numerical examples

In this Section, aforementioned concepts are explained on simple illustrative examples.

Example 1. (trivial) Let all permutations be equal. Then we have: $\pi_i = \pi_m \ \forall i$, $M = 0$, so $d(\pi_m, \pi_r) \leq \frac{2M}{k} = 0$ and $\pi_r = \pi_m$ as expected.

Example 2. Suppose that three DMs rank five alternatives A, B, C, D and E as follows: $DM_1 = \pi_1$ = (A, C, E, B, D), $DM_2 = \pi_2$ = (C, A, B, E, D) and $DM_3 = \pi_3$ = (A, B, C, D, E). We compute the distances among all permutations using the Kendall's tau distance: $d(\pi_1, \pi_2) = 2$, $d(\pi_1, \pi_3) = 3$ and $d(\pi_2, \pi_3) = 2$. Median is π_2 (the problem is well-defined), $M = 4$, $r = \frac{2M}{k} = \frac{8}{3}$, so we look for the consensus up to distance 2 from the median (the distance is an integer). This means only 14 permutations (see Table 1) out of 120 permutations have to be tested. The consensus is (A, C, B, E, D).

Example 3. Now consider different situation: $DM_1 = DM_2$ = (A, C, E, B, D) and $DM_3 = DM_4$ = (A, B, C, D, E). Clearly, we have a situation with two opposing groups, where consensus cannot be achieved. It is easy to verify that the median (as well as consensus) is each of the four permutations, hence the problem is ill-defined.

Example 4. Assume that there are the following preference rankings of five objects: (A, B, C, D, E), (B, A, C, D, E), (B, C, A, D, E), (B, C, D, A, E) and (B, C, D, E, A). All permutations changes by one transposition of A. The median is the 'middle' permutation (B, C, A, D, E), $M = 6$, $r = 12/5$, and $\pi_r = \pi_m$.

6 Computer simulations

To test the behaviour, applicability and efficiency of the reduction technique, extensive Monte Carlo simulations were performed for different values of *n* (number of alternatives) and *k* (number of DMs) with more than 100,000 randomly generated problems. The analysis was performed by the *OCRP Solver*, a decision support tool available from: *http://www.fit.vutbr.cz/~ifiedor/ocrpsolver/*. The consensus was searched by 3 different methods: DCM, CRM (both briefly described in Section 1) and MED, because the consensus found by different methods might differ in general. The MED method is a simple method which finds a consensus as the permutation(s) with the lowest distance to all other permutations given by DMs. The reason to use it is that it searches the consensus in a similar way as the median is searched, so it might be interesting to find out if the consensus found by this method will be close to the median. The results of simulations are shown in Tables 2-5.

Number of alternatives (*n*)	4	5	6	7	8	9	10	12	14	16	18	20
Well-defined problems (expressed in percent)	0.46	27.34	64.84	82.11	87.29	89.9	91.87	94.51	95.58	96.62	97.46	97.96

Table 2 A distribution of well-defined problems in OCRP for different *n* (100 DMs, 10 000 problems).

First, we tested how many problems are well-defined for different *n* and the results are shown in Table 2. As can be seen, for a larger number of alternatives the number of well-defined problems is growing (presumably to 100%). We also examined whether one median implies a unique consensus, but the answer was negative.

Next, we focused on analysing the distance between a median and a consensus for different values of *n*. The results are presented in Table 3. From these results, some observations can be made:

- The average radius $r = \frac{2M}{k}$ was getting larger with growing n (and the space S_n which need to be searched).
- The subspace N of S_n was not much smaller than S_n on average (see the $avg\left[|N|/|S_n|\right]$ values expressing the ratio between the reduced space N and the original space S_n in percent). However, the average (empirical) distance between the median and the furthest consensus (the solution might contain more than one consensus if they are all equally good) was much smaller than r (see the $avg\left[d(\pi_r,\pi_m)/r\right]$ values expressing the ratio between the actual distance $d(\pi_m,\pi_r)$ and the theoretical maximal distance r, as r differs for each problem), which means that in order to find the consensus a significantly smaller subspace of S_n was enough to search through. For example, for n = 8 and the CRM method, the average distance found between median and the furthest consensus was just $0.2 \cdot avg\left[r\right] = 4.458$. It means that, in average, it would be sufficient to process only the permutations with the distance from median up to 4 and there are only 285 such permutations from all 40,320 permutations. That is a difference in more than two orders of magnitude.

Number of alternatives (n)	$avg\left[r\right]$	$avg\left[|N|/|S_n|\right]$	$avg\left[d(\pi_r,\pi_m)/r\right]$					
			π_r nearest from π_m			π_r furthest from π_m		
			CRM	DCM	MED	CRM	DCM	MED
4	4.5362	83.33%	0.044894	0.000986	0.160571	0.103846	0.08047	0.29142
5	7.7154	88.33%	0.066591	0.023921	0.272881	0.117663	0.215662	0.357354
6	11.7322	93.19%	0.100863	0.068657	0.317233	0.128777	0.256401	0.389836
7	16.5156	96.55%	0.134482	0.105069	0.339702	0.160821	0.279646	0.402664
8	22.2904	98.44%	0.178254	0.141813	0.370139	0.202144	0.304321	0.431101

Table 3 Average distance between a median and a consensus for different n (10 DMs, 1000 problems).

Number of alternatives (n)	30% of $avg\left[r\right]$	$avg\left[|N_H|/|S_n|\right]$	Number of solutions containing at least one π_r in N_H		
			CRM	DCM	MED
4	1.36086	16.67%	93.85%	100.0%	77.95%
5	2.31462	11.67%	92.49%	99.34%	60.26%
6	3.51966	6.8%	90.92%	97.38%	47.08%
7	4.95468	3.2%	86.60%	96.78%	39.01%
8	6.68712	3.0%	80.58%	93.61%	30.04%

Table 4 The amount of solutions with at least one consensus in space N_H defined by 30% of r for different n.

Of course, we never know if we have found the consensus until we have processed all the permutations in N, so we cannot search just the permutations up to the average distance. But we can use the reduction technique as a heuristic and try to search only a small part of N in order to find the closest permutation to the preferences given by DMs in a reasonable time, even if we do not know if the permutation is the actual consensus or not. It is still better to find a suboptimal solution to the problem than not find any solution at all. Table 4 shows the probability that we will find at least one consensus (out of possibly many equally good) if we search only a subspace of N up to the 30 % of the average r. As can be seen, the reduction is significantly better in this case as we for, e.g., n = 8 search only 3 % of S_n (compared to 98.44 % before), and we still find a consensus in more that 80 % cases.

Number of DMs (k)	Entropy	$avg\left[r\right]$	$avg\left[d(\pi_r,\pi_m)/r\right]$					
			π_r nearest from π_m			π_r furthest from π_m		
			CRM	DCM	MED	CRM	DCM	MED
10	0.816470	16.51560	0.134482	0.105069	0.339702	0.160821	0.279646	0.402664
5	0.640592	14.60120	0.124172	0.163831	0.319748	0.164463	0.241548	0.740827
3	0.465760	12.18071	0.082859	0.145048	0.285900	0.116300	0.215913	0.966611

Table 5 Average distance between a median and a consensus for different entropy of DMs' preferences (n = 7).

One problem with searching only a small part of N is that with higher n, the average radius r and the distance between the median and consensus is getting higher, decreasing the probability that we will find the consensus in this small part of N. As the problems randomly generated by the Monte Carlo algorithm consist usually from rankings uniformly distributed among the whole space S_n, we are, in fact, dealing with worst case scenarios as real problems normally contain rankings relatively close to each other and not distributed among the whole space S_n. To find out how the average radius r and the distance between the median and consensus will be like when

we are more close to the real problems, we examined the relationship between the entropy of DMs' preferences and the distance between median and consensus. In [6] the entropy of decision makers' preferences was introduced as a measure of concordance among DMs. The lower entropy means larger agreement of DMs' preferences and vice versa. Our simulations showed that when entropy was lower, the radius r was smaller, and also the empirically found distance between median and consensus was lower on average for DCM and CRM, see Table 5. Based on these results, we believe that for real problems the reduction will be much more effective. Moreover, also the number of well-defined problems was getting larger with the decreasing entropy.

To summarize the results, it was learned that the actual distance between a median and consensus was much smaller than the average radius r (from nearly 0 % to 37 % of r for selected k and n) derived from the distance theorem, which means that the consensus often lies in a close neighbourhood of the median and the number of permutations processed can be reduced significantly. Based on these empirical results, the proposed reduction technique can also be considered an efficient heuristic, especially for problems with similar preferences of DMs (with lower entropy), where the radius r is small and the number of permutations, which must be processed, is low. At last, it should be noted that the technique was tested on randomly generated problems, but real problems are far from being random, so the true test of the technique is possible only by practice.

7 Conclusions

The aim of the article was to introduce and test a new optimization technique for the permutation methods for the solution of the ordinal consensus ranking problem. The technique lowers the computational complexity of these methods, which is NP hard due to $n!$ permutations, by reducing the number of permutations they must process in order to find a consensus. It is based on the distance theorem which postulates the maximal theoretical distance of the consensus from one of the initial preferences (called median). Monte Carlo simulations showed that while the maximal theoretical distance was often too large, the empirically found distance between the median and the consensus was much smaller on average. This allows to use this technique also as an efficient heuristic searching only a small part of the reduced space with a high probability to find a consensus, or permutation close to it, in a reasonable time. Moreover, if the decision makers' preferences are similar, the maximal theoretical distance and the empirically found distance are getting lower, making the reduction more significant and beneficial.

Acknowledgements

One of authors (Jiří Mazurek) was supported by the Grant Agency of the Czech Republic (no. 402090405).

References

[1] Arrow, K. A Difficulty in the Concept of Social Welfare. *Journal of Political Economy*, Vol. 58**,** Nr 4, 1950, pp. 328–346.
[2] Borda, J. C. M´emoire sur les ´elections au scrutin. In: *Histoire de l'Acad´emie Royale des Sciences*. Paris, 1781.
[3] Condorcet, M. J. A. N. *Essai sur L´Application de L´Analyse a la Probabilite des Decisions Rendues a la Pluralite des Voix*. Paris, 1785.
[4] Cook, W. D. Distance-Based and Ad Hoc Consensus Models in Ordinal Preference Ranking. *European Journal of Operational Research*, Nr. 172, 2006, pp. 369-385.
[5] Cook, W. D., Kress, M. Ordinal ranking with intensity of preference. *Management Science*, 1985, vol. 31, No. 1, pp. 26-32.
[6] Fiedor, J. Mazurek, J. The Evaluation of the Entropy of Decision Makers' Preferences in Ordinal Consensus Ranking Problem. In: *Proceedings of the 29th International Conference on Mathematical Methods in Economics 2011*. Professional Publishing: Prague 2011, pp. 157-162.
[7] Kendall, M. Rank Correlation Methods, Hafner, New York, NY, USA, 3rd edition, 1962.
[8] Kendall, M., Babington-Smith, B. The Problem of *m* Rankings. *The Annals of Mathematical Statistics*, 1939, Vol. 10, Nr. 3, p. 275–287.
[9] Margolius, B. H. Permutations with Inversions. *Journal of Integer Sequences*, Vol. 4, 2001, pp. 1-12.
[10] Mazurek, J. The Fuzzy Permutation Model for the Ordinal Ranking Problem Solution. In: *Proceedings of the 29th International Conference on Mathematical Methods in Economics 2011*. Professional Publishing: Prague, 2011, pp. 462-467.
[11] Mazurek, J. Group Decision Making in Ordinal Consensus Ranking Problem. In: *Proceedings of the 28th International Conference on Mathematical methods in Economy*, University of South Bohemia, České Budějovice, 2010.
[12] Muir, T. On a simple term of a determinant. *Proc. Royal S. Edinborough*, Vol. 21, 1898, pp. 441-477.
[13] Tavana, M., LoPinto, F., Smither, J. W. A Hybrid Distance-Based Ideal-Seeking Consensus Ranking Model. *Journal of Applied Mathematics and Decision Sciences*, 2007, 18 pages.

Publication Efficiency at DSI FEM CULS – An Application of the Data Envelopment Analysis

Martin Flégl, Helena Brožová [1]

Abstract. The education and research efficiency at universities has always been very important factor. Educational institutions as university departments receive financial resources according their successful performances. In our contribution we compiled the model of the Data Envelopment Analysis for evaluating publication and research efficiency at Department of Systems Engineering, Faculty of Economics and Management, Czech University of Live Sciences Prague in the period 2008-2011. Measured units are department academic staff divided into four categories as Ph.D. students, Lecturer staff + Technical workers, Associate professors and Professors. The attention is also paid to changes in the staff position during the evaluated period. The model outputs are points from publication and research activity. According to the input and output form various versions of the basic model were calculated and the results were analysed.

Keywords: Data Envelopment Analysis, Research and Development results, publication activity, Malmquist index.

JEL Classification: C61, C65
AMS Classification: 90B50, 90C90

1 Introduction

The methods for evaluation of the Research and Development results (R&D) are intensively discussed within the field of scientific policy. The main goal of evaluation is to provide information on research results that were created due to financial support from public resources, and also to gain an insight into the efficiency of such financing. The problem of R&D performance is also discussed at the Czech University of Life Sciences Prague (CULS) and its faculties and departments. The result of this discussion at the Faculty of Economics and Management (FEM) is the Motivation Programme [9] which was introduced in 2010. The program aims are to stimulate publication and research activity of the all academics.

The quantitative evaluation of the organisation has direct implications for financing universities, research organisations and others. From this point of view, the achieved scores indicate the scientific productivity of the organisation. Despite the fact that the official evaluation has many weaknesses, a different tool is not available to enable R&D results to be quantitatively evaluated on the same level of exactness and complexity as the current system.

The official evaluation process is based on formalised procedure which distinguishes between two categories of results [1]:

- Results of basic research – books, papers in scientific journals, conference proceedings;
- Results of applied research – patents, prototypes, industrial designs, maps, certified methods, software.

Each of these results is ascribed a score, such as 20 points for a book, a paper in a journal of the impact factor (IF) receives a score within the interval 10 – 305 (according to the journal ranking), and certified methods approved by a State administration body are valued at 40 points, etc. The evaluation is carried out for each organisation (such as a university), whereby the organisation gains the relative share equal to the share of the authors who created the outcome and are affiliated to the given organisation.

In our contribution we propose to use the model of the Data Envelopment Analysis (DEA) for evaluating publication and research efficiency and we show its application at Department of Systems Engineering, Faculty of Economics and Management, Czech University of Live Sciences Prague (DSI FEM CULS) in the period

[1] Czech University of Live Sciences in Prague, Faculty of Economics and Management, Department of Systems Engineering, Kamycka 129, 165 21 Praha 6 – Suchdol, flegl@pef.czu.cz, brozova@pef.czu.cz.

2008-2011. Academic staff of department is divided into four categories as Ph.D. students, lecturer staff + technical workers, associate professors and professors are Decision Making Units (DMUs). The attention is also paid to changes in the staff position during the evaluated period. The model outputs are points from publication and research activity. DEA models for three periods and the Malmquist index are calculated. The results of the models are also interpreted in a graphical form.

2 Materials and Methods

2.1 Efficiency measuring in educational institutions

The reliability of the results depends on the accurate selection of the data best adapted to the objective of the study. Bessent et al [2] pointed out the major problems in educational efficiency measurement which are also important in process of R&D evaluation:

- Obtaining data to specify adequate input measures;
- Obtaining data to specify outputs that were not limited to cognitive test results;
- Difficulties in communicating the results to those affected by them.

It is possible to find many scientific studies based on measuring efficiency in educational environmental. Worthington [11] summarises the approaches which have been used for measuring efficiency in educational institutions (high schools, universities, study programme efficiency, etc.) between 1981 and 1998. The DEA method was the dominant method of the educational efficiency measurement. These DEA models mainly contain the number of teaching, administrative and support staff as the inputs.

The outputs are thus the papers and letters in academic journals, authored/edited books, published works [6]. Kao and Hung [7] compiled model for measuring department efficiency. Model contains personnel, operating expenses and floor space as the inputs. The outputs were total credit hours, publications and external grants. Jablonský [4] presented the DEA model for measuring resources' allocation among university departments. The model contains the number of hours of direct and indirect teaching and the quality of research as the outputs. DEA models for measuring departmental efficiency can also be focused on improving teaching performance. Montoneri et al [8] used the richness of course contents and the diversity of accessed multiple teaching channels as the inputs. The outputs were thus the positive degree of teaching attitude and students' learning performance.

2.2 Data Envelopment Analysis Method

DEA evaluates DMUs against the best DMUs with the idea: if some DMU can produce a certain level of output utilizing certain level of input, another DMU of equal scale should be capable of doing relatively the same. DEA is a nonlinear programming model for the estimation of productive efficiency of DMUs based on relationship between multiple outputs and multiple inputs. These outputs and inputs are usually of various characters and of variety of forms which are difficult to measure. The DEA measure of the efficiency of any DMU is obtained as the maximum of a ratio of weighted outputs to weighted inputs subject to the condition that the similar ratio for every DMU is less than or equal to 1.

The simplest DEA model assumes constant returns to scale, this model is called CCR model according to its authors Charnes, Cooper, and Rhodes [3]. Let y_{jk} be the amount of the j^{th} output from unit k, and x_{ik} be the amount of the i^{th} input to the k^{th} unit. Using the CCR model the DMU efficiency of a particular unit H is calculated using the following linearization of original DEA model. Primal and dual CCR output oriented models are formulated as:

Primal model

$$\Phi_H = \sum_{i=1}^{m} v_{iH} x_{iH} \rightarrow MIN \tag{1}$$

$$\sum_{j=1}^{n} u_{jH} y_{jH} = 1$$

subject to

$$\sum_{i=1}^{m} v_{iH} x_{ik} - \sum_{j=1}^{n} u_{jH} y_{jk} \geq 0, \quad k = 1, \ 2, \ldots, p \tag{2}$$

$$u_{jH} \geq 0, \ j = 1, \ 2, \ldots, \ n \ \ a \ \ v_{iH} \geq 0, \ i = 1, \ 2, \ldots, \ m.$$

Dual model

$$z_H \rightarrow MAX \tag{3}$$

subject to

$$\sum_{k=1}^{p} \lambda_{kH} x_{ik} \leq x_{iH},\ i = 1,\ 2,\ldots,\ m$$

$$z_H y_{jH} - \sum_{k=1}^{p} \lambda_{kH} y_{jk} \leq 0,\ j = 1,\ 2,\ldots,\ n \tag{4}$$

$$\lambda_{kH} \geq 0,\ k = 1,\ 2,\ldots,\ p$$

The decision variables $u = (u_1,\ldots,u_m)$ and $v = (v_1,\ldots,v_n)$ are the weights given to the m outputs and to the n inputs respectively. To obtain the relative efficiencies of the all units, the model is solved for one unit at a time. The decision variables $\lambda = (\lambda_1,\ldots,\lambda_p)$ are the weights given to the efficient DMUs for creating virtual (efficient) DMU corresponding to non efficient DMU. The inputs and outputs of virtual DMU are calculated used the formulas:

$$x'_{iH} = \Phi_H x_{iH} - s_{iH}^{-},\ i = 1,\ldots,m$$
$$y'_{jH} = y_{jH} + s_{jH}^{+},\ j = 1,\ldots,n \tag{5}$$

or

$$x'_{iH} = \sum_{k=1}^{p} \lambda_{kH} x_{ik},\ i = 1,\ldots,m$$

$$y'_{jH} = \sum_{k=1}^{p} \lambda_{kH} y_{jk},\ j = 1,\ldots,n \tag{6}$$

where s_{iH}^{-} and s_{jH}^{+} are slacks in the dual constraints.

The constant returns to scale describes the individual constant ability of publication and research work. Output orientation of the model means that results explicitly show the necessary augmentation of outputs with the same amount of inputs. This model orientation reflects the possibility of DMUs to improve his/her research activity.

Analysis of changes of DMU's efficiency over time is based on the Malmquist index [4], which can be used for the investigation of the causes of efficiency change. Malmquist index is defined with constant returns to scale, which allows supposing the same technology in both periods. This convention enables Malmquist index with output orientation quantifies change of efficiency of DMU between period *t* and period *t*+1 and can be formulated as follows

$$M(x^t, y^t, x^{t+1}, y^{t+1}) = \left(\frac{D^t(x^t, y^t) D^t(x^{t+1}, y^{t+1})}{D^{t+1}(x^t, y^t) D^{t+1}(x^{t+1}, y^{t+1})} \right) \tag{7}$$

where $D^t(x^t, y^t)$ is efficiency in the period *t* and $D^{t+1}(x^{t+1}, y^{t+1})$ efficiency in period *t*+1,

$D^t(x^{t+1}, y^{t+1})$ is efficiency in the period *t*+1 considering efficiency frontier in period *t* and

$D^{t+1}(x^t, y^t)$ is efficiency in period *t* considering efficiency frontier in period *t*+1.

Malmquist index greater than 1 indicates productivity gain; Malmquist index less tahn 1 indicates productivity loss; and Malmquist index equal to 1 means no change in productivity from time *t* to *t* + 1.

Authors used Efficiency Measurement System (EMS) software for calculation DEA model [10].

2.3 Model Data

This study fructifies the secondary data from Rejstřík informací o výsledcích/Information Register of R&D results (RIV), which is the key database for the evaluation of scientific work in the Czech Republic. The evaluation is carried out by the Rada pro výzkum, vývoj a inovace/Research, Development and Innovation Council (RVVI). All the results are evaluated by the Metodika hodnocení výsledků výzkumných organizací/Methods for evaluating R&D results [1] which are focused on results that were produced by each research organisation in the last five years. The study is based on the most up-to-date files that refer to R&D results published between 2006 and 2010. These results were officially published by the RVVI in January 2012.

Background data for the DEA model contains DMUs, inputs and outputs (**Table 1**).

DMU	Input	Output (2008)	Output (2009)	Output (2010)	Output (2011)
P1	8	3.153	0	0	30.253
P2	12	8.536	8.278	8.518	28.333
P3	12	18.920	0	5.845	31.666
P4	16	17.738	0	0	5
P5	8	21.286	0	15.305	24.334
P6	8	21.286	0	15.305	15
P7	12	8.536	7.096	0	19.252

Table 1 Input data for DEA model

Evaluated DMUs are expressed for 7 employees of Department of Systems Engineering who published at least one paper in 2008-2009 and 2010-2011 periods. This constraint is important for possibility of DEA model application and for Malmquist index calculation.

Each DMU has only one input expressing the position at the department, i.e. Ph.D. student express one point per year, Lecturer staff + Technical workers two points, Associated professors three points and finally Professors four points for each year. The data were obtained from the university's databases during the period of 2008 - 2011. Data had to be cleaned from imprecise data to guarantee accurate results. We noted changes of each DMU, e.g. if a Ph.D. student graduated in 2009 we must calculate its points as follows: first two years 2 points (one point per year) and from years 2010 and 2011 4 points (two pints per year). From this point of view, the position of authors is precisely classified to a year in which the person published his/her paper. Inputs are calculated by position-year measure with regard to four categories.

The outputs of our DEA model represent the evaluation of publication activities divided into the four categories: publications in 2008, 2009, 2010 and 2011. The points evaluation represents received points from publication/research activities during this period. This publication/research activity was evaluated by RVVI [1]. Points were then summarized for each author and its position with regard to the published year. In the case of more than two authors of a paper, we divided the points in proportion.

3 Results and Discussion

In our contribution, the output-oriented DEA model is used. The reason for output orientation is because the authors want to evaluate publication/research activities. The results will give us information as to who is efficient and who is not. The recommendation for an inefficient employee is going to be an improvement of the publication/research activity. It is also necessary to mention that this model is calculated for the 2008-2011 period.

Table 2 summarizes the efficiency results of all seven employees. In the first column Output Score 2008-2011 shows the efficiency of the DSI's employees. Efficient employees are P1, P2, P5 and P6 which have the efficiency result of 100 %. The column Output score I shows the efficiency result in 2008-2009 period. The efficient DMUs are P2, P5 and P6. The other DMUs are inefficient although the DMU P7 is closed to the efficient line. The results for the same DMUs but for the 2010-2011 period are shown in column Output score II. In this period the group of the efficient DMUs partly changed. P5 and P6 are still efficient. P1 became efficient instead of the P2 which is now inefficient. P2 and P3 are relatively closed to the efficient line.

The main objective of our contribution was to calculate Malmquist index and measure the efficient development during the 2008-2011 period. Malmquist index values are in the last column in the **Table 2**. The DMUs with the Malmquist index greater than 1 have increased their efficiency from the first period to the second period. The highest improvement from measured DMUs reached the DMU P1. Malmquist index value 5,724 means efficiency increased almost 6 times, this employee was really inefficient in the first period and strongly increased his efficiency in the second period. P2, P3 and P7 were almost inefficient in the both periods; nevertheless their efficiency was increasing in the group of analysed DMUs. On the other side those DMUs which have the Malmquist index lower than 1, have decreased their efficiency in monitored periods. This is only a situation of P4. The DMUs which have the Malmquist index equal to 1 are efficient in both periods.

DMU	Output Score (2008-2011)	Output Score I (2008-2009)	Output Score II (2010-2011)	Malmquist index
P1	100.00%	675.02%	100.00%	5.724
P2	100.00%	100.00%	143.48%	1.369
P3	124.00%	168.75%	133.76%	1.237
P4	240.01%	240.01%	1210.12%	0.394
P5	100.00%	100.00%	100.00%	1
P6	100.00%	100.00%	100.00%	1
P7	111.69%	111.69%	235.71%	1.129

Table 2 DMUs' efficiency and Malmquist index

The **Figure 1** describes the DMUs' efficiency changes during the measured periods. Figure contains two efficiency frontiers, the efficiency frontier I for the period 2008-2009 and then the efficiency frontier II for period 2010-2011. Each employee is described by the dotted line; its initial point shows the position (efficiency) in the period 2008-2009, its ending point with arrow represents the position in the second period 2010-2011.

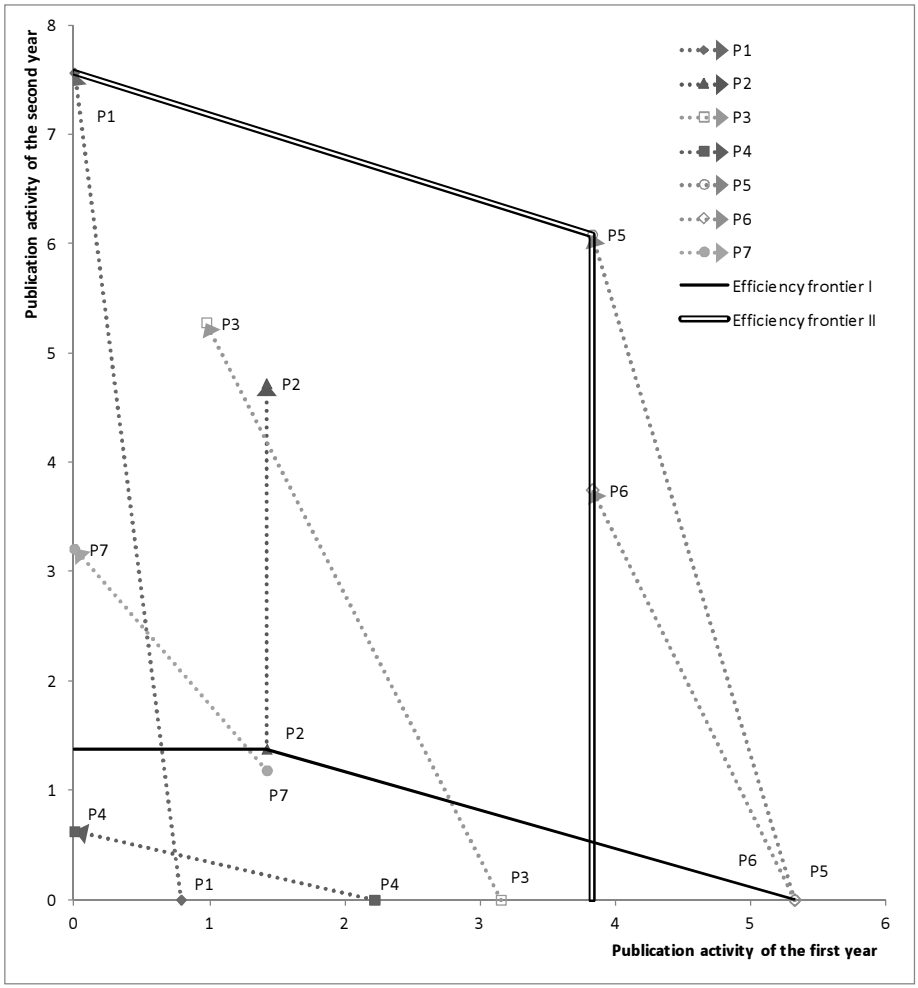


Figure 1 Changes of DMUs' efficiency during the measured periods

According to the DEA methodology all those DMUs which are lying on the frontier are efficient. The DMUs P5 and P6 are efficient in both periods, both DMUs lie on the efficiency frontiers. The unit P6 however is dominated by P5, because P5 is better in the second period, P5 has more publications then P6. The representation of units P3, P4 and P7 explains inefficiency of these units, but also the differences among them. Unit P4 shows significant deterioration of its efficiency, because the ending point of its line is very far from the efficiency frontier of the second period. On the other hand unit P1 is shifted from inefficient position to the efficient point on the efficiency frontier of the second period. And unit P2 is shifted from the efficient point on the efficiency frontier of the first period to inefficient position in the second period.

4 Conclusion

Presented DEA analysis of the efficiency of R&D publication activity of employees shows reasonability of this approach even though the number of analysed DMUs in this case is small.

The important requirement of this approach is simple structure of inputs and outputs system allowing two-dimensional description of the results.

- Inputs data describes category of employee at a department during two-year period;
- Output data specification is based on RIV evidence and analysed period can be only two-year long;
- Graphical representation is the best way how to explain the results of the DEA model.

The efficiency score indicates how the employees have to improve their research activity. These results depend on the selected radial measure, so it is possible for some employees improve their publication activity in other proportion. Malmquist indices calculated for analysed employees show individual position at a department and show the changes of their efficiency. Generally it is not possible to expect that employees will be efficient in following periods. This is caused by the character of research work and long time of research finalisation.

Acknowledgements

This work was supported by the grant project of the Grant Agency of CULS Prague, No. 201221029, "Využití metody DEA pro alokaci veřejných zdrojů mezi produkční jednotky".

References

[1] RVVI: Metodika hodnocení výsledků výzkumných organizací/Methods for evaluating R&D results. [Online], Available: http://www.vyzkum.cz [12 April 2012], 2010.
[2] Bessent, A., Bessent, W., Kennington, J., Reagan, B.: An application of mathematical programming to assess productivity in the Houston independent school district. *Management Science*, 28, pp. 1355-1367, 1982
[3] Charnes, A., Cooper, W., Rhodes, E.: Measuring the efficiency of decision-making units. *European Journal of Operational Research,* vol. 2, pp. 429–444, 1987.
[4] Cooper, W., Seiford, L. M., Tone, K.: *Data Envelopment Analysis.* 1st ed. New York: Springer Publisher, 2006.
[5] Jablonský, J.: Models for efficiency evaluation in education. In: *Proceedings of the 8th International Conference on Efficiency and Responsibility in Education (ERIE 2011),* Prague, pp. 88-97, 2011.
[6] Johnes, J., Johnes, G.: Research funding and performance in U.K. university departments of economics: A frontier analysis. *Economics of Education Review*, 14, pp. 301-314, 1995.
[7] Kao, Ch., Hung, H.T.: Efficiency analysis of university departments: An empirical study. *Omega*, 36, pp. 653-664, 2006.
[8] Montoneri, B., Lin, T. T., Lee, Ch. Ch., Huang, S.-L.: Application of data envelopment analysis on the indicators contributing to learning and teaching performance. *Teaching and Teacher Education*, 28, pp. 382-395, 2012.
[9] Motivační program pro pracovníky a doktorandy PEF ČZU v Praze v roce 2012/Incentive program for employees and postgraduates of FEM CULS in Prague in 2012. Pef ČZU, Praha, 2012.
[10] Scheel, H.: *EMS: Efficiency Measurement System*. [Online], Available: http://www.holger-scheel.de/ems/ [12 April 2012], 2000.
[11] Worthington, A.: An Empirical Survey of Frontier Efficiency Measurement Techniques in Education. *Education Economics*, 9 (3), pp. 245-268, 2001.

Monetary policy effects: comparing macroeconomic impulse responses for the Visegrad Group countries

Tomáš Formánek[1], Roman Hušek[2]

Abstract. Our paper focuses on the analysis and comparison of monetary policy impacts on industrial production and inflation in different EU countries. The Visegrad Group (VG) countries (Czech Republic, Slovak Republic, Poland and Hungary) and their specific impulse response dynamic behaviours are compared against each other and with selected euro zone countries. This comparison is based on a standardized VAR model specification (using monthly data from 2000 to 2011 for the underlying estimation). We use Blanchard-Quah [1] decomposition scheme to isolate macroeconomic shocks for further analysis. Alternative VAR model specifications may be used to generate impulse response functions with varied degrees of stratification and interpretation possibilities. We evaluate the homogeneity of expected costs of disinflationary monetary policy among VG countries by studying the impulse response dynamics of the output cost of disinflation. Our results indicate that VG countries do not exhibit a statistically significant symmetric behaviour in response to incurred monetary shocks. The nature of heterogeneous behaviour found is discussed and results are compared to previous studies with similar focus.

Keywords: VAR, impulse-response, monetary policy.

JEL Classification: C32, E31, E52
AMS Classification: 91B64

1 Introduction

Regardless of the general operational framework or individual monetary policy measures applied by the European Central Bank (ECB), individual euro area economies should exhibit reasonably homogenous responses to monetary shocks. Otherwise, conducting common monetary policy would be politically unsustainable in the long term. For example, as the timing and extent of a disinflationary policy action is decided by the ECB upon weighting the expected benefits and costs of such action, countries prone to unfavourable asymmetric macroeconomic shock/response behaviour might soon perceive such restrictive measures as being carried out at their expense. Hence, possible heterogeneities in different aspects of macroeconomic dynamics among prospective euro area member states were widely addressed even prior to establishing the Economic and Monetary Union (EMU). Demertzis et. al. [4] provide relevant overview in their study. Also, ever since the euro currency was actually introduced, there has been an increased interest in studying the potential cross-country differences in monetary policy transmission mechanisms (see e.g. Jarociński [9] for a review of recent literature).

We compare the impulse-response (IR) dynamics (see Hušek [6] for description) relevant for sacrifice ratio (SR) estimates among different EU economies, with main focus on Visegrad Group (VG) countries. Our data panel consists of 2000 – 2011 monthly variable observations for Czech Republic, Slovakia, Hungary, Poland, Austria, Germany, France and EMU total. This article is structured as follows: Next section describes the theory behind IR dynamics used for our calculation and provides a brief literature overview. Third section contains the specification of our econometric model and data used for estimation, along with results and their detailed interpretation. Last section and the list of references conclude our article.

2 The IR dynamics of output cost of disinflation

Evaluation of the expected costs of restrictive (disinflationary) monetary policy is usually based on the SR coefficient, which measures the output cost of disinflation. As in Cecchetti and Rich [2], SR is defined as the cumulative loss in relative GDP growth associated with a one percentage point of permanent reduction in the inflation rate. SR also closely relates to inflation targeting, as we may use it as a proxy for nominal inertia of the economy. Higher SR values indicate a less favourable trade-off between price stability and GDP growth.

Unfortunately, individual SR estimations based on real (observed) macroeconomic data are notorious for contradicting conclusions reached by different authors (see e.g. Cecchetti and Rich [2] or Jarociński [9] for ex-

[1] University of Economics, Prague, Dept. of Econometrics, nám. W. Churchilla 4, Praha 3, formanek@vse.cz.
[2] University of Economics, Prague, Dept. of Econometrics, nám. W. Churchilla 4, Praha 3, husek@vse.cz.

tensive literature review). Therefore, due to limitations in available space, we focus mainly on evaluation of the **homogeneity of expected reactions** to disinflationary monetary policy actions. By assessing the homogeneity of impulse response dynamics of the output cost of disinflation for different economies, we effectively approach this topic from the **optimum currency area** (OCA) perspective, as described in detail by Demertzis et. al. [4].

For industrialized countries, it is common to use the loss of industrial production growth instead of the GDP growth for SR analysis, e.g. as shown in Corbo et. al. [3] or Jarociński [9] and we adopt a similar approach in this contribution. As in Hušek and Formánek [7], it should be mentioned that our model is built on the implicit assumption that effects of monetary restrictions and expansions on industrial production, inflation and real interest rate are not systematically different, except for their sign (direction of the effect).

The use of Blanchard-Quah [1] **BQ decomposition** in this contribution may be justified using the following reasoning: as we analyze actual (observed) macroeconomic time series for different countries, we cannot attribute all fluctuations in industrial production, inflation and real interest rate to monetary policy actions. As a first step, we need to identify and isolate monetary policy shocks from other types of shocks. BQ decomposition is a method for identifying the unobservable (e.g. monetary policy) orthogonalized (mutually independent) shocks from residuals obtained by estimating certain types of VAR models. BQ decomposition is based on additional long term economic restrictions in mathematical form. For a simplistic example of BQ decomposition usage, let's assume a VAR model containing only 2 variables: real GDP growth and inflation. We may use IR dynamics to simulate neoclassical supply and demand shocks by implementing an additional assumption that demand shocks in the long term affect only inflation, not real GDP (that is the BQ zero-value identifying restriction). Cecchetti and Rich [2] provide references to literature covering BQ decomposition theory and applications.

3 Econometric modelling and empirical results

Our baseline model estimations for subsequent IR-dynamics comparison is based on a reduced second order VAR specification, VAR(2). For each country *c*, the model may be written as

$$\boldsymbol{y}_{ct} = \boldsymbol{A}_{c1}\boldsymbol{y}_{ct-1} + \boldsymbol{A}_{c2}\boldsymbol{y}_{ct-2} + \boldsymbol{B}_{c}\boldsymbol{x}_{ct} + \boldsymbol{u}_{ct} \quad , \tag{1}$$

where $\boldsymbol{y}_{ct}$ is a 3x1 vector of endogenous variables (industrial production index, CPI inflation, real interest rate) for country *c*, matrices $\boldsymbol{A}_{c1}$ and $\boldsymbol{A}_{c2}$ contain 3x3 coefficients for the first and second lags of vector $\boldsymbol{y}_{ct}$. $\boldsymbol{x}_{ct}$ is a 5x1 vector of exogenous elements (first and second lag of country-specific nominal effective exchange rate, country specific constant term, common oil price inflation, common dummy variable reflecting the 2008 crisis onset), $\boldsymbol{B}_{c}$ is a 3x5 matrix of coefficients for $\boldsymbol{x}_{ct}$ and the 3x1 vector $\boldsymbol{u}_{ct}$ contains *random innovations with a potential for contemporaneous correlations*, as shown e.g. in Hušek [6].

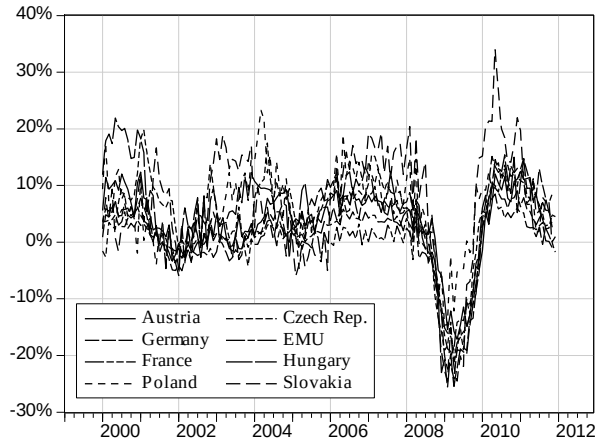


Figure 1 Industrial production: y-o-y index

Model (1) for all countries is estimated using monthly data (2000M03 to 2011M12) from the International Monetary Fund's IFS database (available from http://elibrary-data.imf.org/). Industrial production index (real values, 2005 prices) and CPI inflation are calculated as y-o-y relative changes and real interest rate is the ex-post real p.a. yield from short term assets (nominal interest rate adjusted by CPI index). We follow Jarociński [9] to some extent by introducing the nominal effective exchange rate (NEER) y-o-y relative change, in order to eliminate *price puzzle* IR behaviour, i.e. price level increases in response to a simulated restrictive monetary shock.

World factors are represented by oil price inflation (y-o-y index). From Figure 1, we may observe the onset of the 2008 crisis on industrial production. In order to control for this important shock, we introduce a dummy variable covering twelve consecutive months starting from 2008M09. For the sake of consistency, common dummy variable specification is used for all economies considered, although statistically significant country-specific dummies can be introduced and justified by data mining methods. Similarly, unified VAR(2) specification was also chosen with respect to consistency of estimates throughout different economies. The use of y-o-y relative indexes allows for direct comparison among different countries, eliminating the need for further data transformation into a common currency denomination. EViews 6 software was used for all estimations.

Due to limited space available, we only show results for Czech Republic and Slovakia in Table 1. The remaining estimates, econometric verification outputs and IR-dynamics data are available from the authors upon request. Along with other econometric verification tests, the R^2 coefficients and F-tests provide sufficient assurance for the next step: computation of IR dynamics for evaluation of macroeconomic dynamics homogeneity. Unlike Jarociński [9], we do not compare the individual regression coefficients of the estimated model (1) among different economies, as observed (real) variables are more susceptible to multicolinearity issues than carefully curated artificial sampling. Instead, we exercise the OCA approach by pair wise comparison of the responses of industrial production index and inflation to ε_{ct}^{LM} (monetary shock) among the different economies.

Once stationarity is tested both for data up to 2008M09 (pre-crisis) and for residuals of the model (1) estimated after controlling for the crisis-related fluctuations using the dummy variable previously described, we may re-write our VAR(2) model by switching to a vector moving average (VMA) representation:

$$\begin{bmatrix} ip_{ct} \\ \pi_{ct} \\ (i_{ct} - \pi_{ct}) \end{bmatrix} = \boldsymbol{A}(L) \begin{bmatrix} \varepsilon_{ct}^{ip} \\ \varepsilon_{ct}^{LM} \\ \varepsilon_{ct}^{IS} \end{bmatrix} . \tag{2}$$

In the VMA specification (2) as pioneered by Shapiro and Watson [11], ip_{ct}, π_{ct} and (i_{ct} - π_{ct}) are endogenous variables of the model (1) and $\boldsymbol{A}(L)$ is a 3x3 matrix of polynomial lag operators, as described e.g. in Cecchetti and Rich [2]. As shown by Shapiro and Watson [11], random elements may be interpreted as follows: ε_{ct}^{ip} is an aggregate supply shock, ε_{ct}^{LM} and ε_{ct}^{IS} are aggregate demand shocks (stratified into IS and LM shocks). Hence, for our calculation we associate the non-systematic changes in monetary policy with LM shock, i.e. with ε_{ct}^{LM}. Further theoretical background for such association is provided e.g. by Hušek and Formánek [7]. Estimates of the VAR(2) model from equation (1) were consistently performed for each country in our panel.

BQ decomposition is used to identify the *unobservable orthogonalized shocks* ε_{ct}^{ip}, ε_{ct}^{LM} and ε_{ct}^{IS} from equation (2). Using theoretical reasoning provided by Cecchetti and Rich [2], the additional BQ long-term zero value identifying conditions hold for coefficients c_{12}, c_{13} and c_{23} of the 3x3 long-term BQ identification matrix $\boldsymbol{C}_c$ containing long run cumulative responses of endogenous variables (in rows) to structural shocks (in columns) for each country c. Therefore, demand shocks ε_{ct}^{LM} and ε_{ct}^{IS} have no long-term cumulative effect on industrial production ip_{ct} and IS shock ε_{ct}^{IS} has zero long term cumulative effect on inflation π_{ct}.

Interpretation of empirical results

In this paper, we explore the IR dynamics directly related to SR calculation, i.e. the responses of ip_{ct} and π_{ct} to simulated LM shocks. At the same time, a few related findings should be briefly mentioned: Given Germany's leading economic role in the region, we have experimented with adding lagged German industrial production index as an exogenous variable to VAR(2) models for the remaining individual economies, aiming for better distinction between domestic dynamics and external developments. This approach provides statistically sound and theoretically relevant results for individual economies, yet due to inconsistent statistical significance of lags of German ip_{ct} used as exogenous variable throughout other economies and for the sake of available space we decided to drop such specification from this contribution. Also, from Figure 1 we may see that Slovakia has experienced the most severe relative y-o-y ip_{ct} fluctuations in the observed time series of all the VG countries: -24.5 % in 2009M02, 2009M05 and +34.0 % in 2010M05. Because Slovakia's real exports and other publicly available data for the 2008-2010 period also exhibit fluctuations exceeding other VG countries, we plan to further econometrically approach this topic in order to asses weights of possible causes for this behaviour: absence of monetary policy tools due to euro accession, higher fiscal deficits, decline in foreign direct investment (FDI) dynamics, etc.

Vector Autoregression Estimates. Sample: 2000M03 2011M12. Included observations: 142 (after adj.) Standard errors in ( ).						
	Czech Republic			**Slovakia**		
	ip_{ct}	π_{ct}	$(i_{ct} - \pi_{ct})$	ip_{ct}	π_{ct}	$(i_{ct} - \pi_{ct})$
$ip_{c\ t-1}$	0.598	-0.010	0.016	0.489	0.011	-0.011
	(0.087)	(0.013)	(0.013)	(0.087)	(0.014)	(0.014)
$ip_{c\ t-2}$	0.142	0.017	-0.017	0.277	0.006	-0.005
	(0.082)	(0.012)	(0.012)	(0.079)	(0.013)	(0.013)
$\pi_{c\ t-1}$	2.619	1.194	-0.035	0.608	1.472	-0.157
	(1.977)	(0.290)	(0.288)	(2.121)	(0.340)	(0.335)
$\pi_{c\ t-2}$	-2.682	-0.227	0.062	-0.655	-0.549	0.212
	(1.984)	(0.291)	(0.289)	(2.086)	(0.334)	(0.330)
$(i_{c\ t-1} - \pi_{c\ t-1})$	3.803	0.134	0.892	0.345	0.434	0.886
	(1.991)	(0.292)	(0.290)	(2.177)	(0.349)	(0.344)
$(i_{c\ t-2} - \pi_{c\ t-2})$	-3.558	-0.124	0.070	0.060	-0.433	0.082
	(1.928)	(0.283)	(0.281)	(2.120)	(0.340)	(0.335)
Constant	2.048	0.081	-0.099	1.874	0.173	-0.084
	(0.771)	(0.113)	(0.112)	(2.091)	(0.335)	(0.330)
Dummy variable (*t*)	-6.235	-0.240	0.277	-7.784	0.145	-0.086
	(1.749)	(0.257)	(0.255)	(2.233)	(0.358)	(0.353)
Oil price inflation (*t*)	0.014	0.003	-0.003	0.008	0.002	-0.002
	(0.009)	(0.001)	(0.001)	(0.015)	(0.002)	(0.002)
NEER_index (*t*-1)	-0.214	-0.013	0.010	0.260	-0.006	-0.004
	(0.138)	(0.020)	(0.020)	(0.246)	(0.039)	(0.039)
NEER_index (*t*-2)	0.129	-0.008	0.007	-0.191	-0.005	0.015
	(0.131)	(0.019)	(0.019)	(0.247)	(0.040)	(0.039)
R-squared	0.849	0.945	0.903	0.807	0.951	0.928
Adj. R-squared	0.837	0.941	0.896	0.792	0.947	0.923
S.E. equation	3.132	0.459	0.456	4.778	0.766	0.755
F-statistic	73.504	226.997	121.865	54.760	252.905	169.715
Akaike AIC	5.195	1.356	1.342	6.040	2.378	2.351

Table 1 VAR(2) model estimates for Czech Republic and Slovakia

Based on the estimated model (1) and using BQ decomposition, we may now generate a set of 72 IR functions (responses of 3 endogenous variables to each of the 3 orthogonalized innovations, for all 8 economies) with each IR series spanning 5 years (60 months). For this contribution, we only follow up on the responses of ip_{ct} and π_{ct} to a simulated restrictive monetary shock ε_{ct}^{LM}. While the introduction of lagged NEER$_c$ explanatory variables to our model has improved the overall IR dynamics attributes, it was not completely successful in eliminating the price puzzle. Poland still exhibits sustained price puzzle behaviour and for Slovakia and Germany we experience brief π_{ct} increases (oscillations) after the simulated negative monetary shock ε_{ct}^{LM}.

	Austria	**CZ**	**Germany**	**Slovakia**	**EMU**	**France**	**Hungary**	**Poland**
Austria	1.0000	0.8079	0.4615	-0.0846	0.9051*	0.6091	0.9797*	-0.2304
CZ	0.8116	1.0000	0.8220	0.3965	0.6938	0.3902	0.7910	-0.5991
Germany	0.9228*	0.7970	1.0000	0.7699	0.1983	-0.1773	0.3736	-0.9347
Slovakia	0.9628*	0.9255*	0.9393*	1.0000	-0.2869	-0.4688	-0.1601	-0.8966
EMU	-0.0434	-0.1092	-0.4175	-0.1625	1.0000	0.8791*	0.9692*	0.0665
France	-0.8556	-0.7525	-0.9788	-0.8935	0.5375	1.0000	0.7473	0.4237
Hungary	-0.7634	-0.9745	-0.6782	-0.8633	-0.1005	0.6044	1.0000	-0.1176
Poland	-0.8696	-0.9725	-0.7866	-0.9405	-0.0303	0.7358	0.9644*	1.0000

Table 2 Correlations of IR dynamics of ip_{ct} (lower triangle) and π_{ct} (upper triangle) after ε_{ct}^{LM} shock

The responses of ip_{ct} and π_{ct} to simulated restrictive monetary shocks may be efficiently compared using correlation coefficients. Table 2 shows both ip_{ct} and π_{ct} IR dynamics correlation coefficients. ip_{ct} coefficients are in

the lower triangle (under the diagonal) and homogeneity of π_{ct} IR dynamics is evaluated in the upper triangle. So, for Austria and Czech Republic, the coefficient 0.8116 in Table 2 describes the correlation of ip_{ct} reactions to the ε_{ct}^{LM} shock and the coefficient 0.8079 refers to π_{ct} IR dynamics correlation. Due to data limitations, we were unable to eliminate duplicity in correlation coefficients calculated for EMU total. The bias incurred for Austria and Slovakia is negligible, but EMU's IR dynamics correlations with Germany and France should be interpreted with caution. Fisher's *z*-transformation is used for the evaluation of homogeneity of internal macroeconomic dynamics as measured by the IR functions in Table 2: correlation values significantly exceeding an arbitrarily set homogeneity threshold of +0.8 at the 5 % significance level are marked by *. Interestingly, all ip_{ct} IR functions regarded as homogenous at this significance level also hold their homogeneity even at the 1 % level. For illustration purposes, IR dynamics of ip_{ct} for a group of four most tightly positively correlated economies (as per Table 2) are provided in Figure 2.

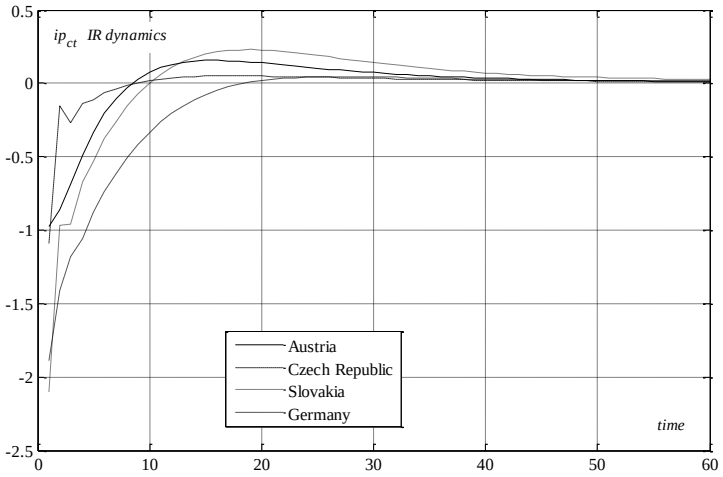


Figure 2 IR dynamics of ip_{ct} resulting from standardized restrictive monetary shock ε_{ct}^{LM}

Correlation coefficients are rather self-explanatory and the level of heterogeneity observed may seem alarming (this interpretation does not change even if the homogeneity threshold is lowered to, say, +0.5). However, the values from Table 2 need to be interpreted considering the theoretical context provided by Mundell [10] and the actual (and arguably successful) existence of common European currency. Also, we should keep in mind that the current euro currency problems are mostly due to prolonged fiscal irresponsibility in certain EMU member states and not induced by asymmetric reactions to monetary shocks.

Hence, in context with previous estimates, as for example in Demertzis et. al. [4] or Hušek and Formánek [8], it is possible to state the following: We do not find a homogenous (common) macroeconomic IR dynamics behaviour pattern among VG countries. Instead, π_{ct} IR functions in Table 2 are rather heterogeneous and for the IR dynamics of ip_{ct} we may discern two main patterns: Industrial production in Czech Republic and Slovakia reacts very similarly to monetary shocks, yet we must bear in mind that only Czech Republic may exercise autonomous monetary policy. Hungary and Poland are also strongly correlated, but they differ from Czech Republic and Slovakia. Although strong historical ties provide a simple explanation for the correlation among ip_{ct} IR dynamics of Czech Republic and Slovakia, it is more difficult to find convincing evidence to support the observed similarity between Hungary and Poland (relatively similar FDI dynamics during the EU accession period era and – generally speaking – ongoing structural changes induced by EU membership should be mentioned as possible significant factors).

Although strong overall heterogeneity of IR dynamics for the economies analyzed was identified, we find evidence for various subgroups of relatively similar economies (i.e. homogeneous IR dynamics). In addition to the two groups mentioned in previous paragraph, Austria-Germany is the most prominent, historically and economically justifiable subgroup. Austria's IR behaviour is also very similar to Czech Republic and Slovakia. On the other hand, Germany and France exhibit strongly dissimilar reactions to monetary shocks, with negatively correlated IR dynamics. The overall heterogeneous IR dynamics among the economies compared should not come as a surprise, they match the findings in other previously published studies on IR dynamics (where GDP growth was used instead of ip_{ct}): For the period 1961Q3 to 1995Q1, Demertzis et. al. [4] have also identified strong asymmetry between Germany and France, as well as among many other 'old' EU economies (EMU candidates at the time of publication of their paper). Hušek and Formánek [8] reach similar conclusions for a different set of EU members, using data from 1997Q2 to 2004Q2. Therefore, robust evidence exists in the analyzed

economies for prevalent heterogeneity in internal macroeconomics dynamics induced by monetary restrictive shocks, as measured using the methodology described here.

4 Conclusions

Our results indicate that Visegrad Group countries do not exhibit a unique symmetric behaviour in response to incurred monetary shocks. The reactions of CPI inflation to a simulated monetary restrictive shock are rather heterogeneous. We find two distinctive patterns of IR dynamics in industrial production: after a monetary shock, the response of industrial production in Czech Republic closely resembles Slovakia and Hungary's IR dynamics is similar to Poland – yet the two subgroups are different from each other.

All economies analyzed in this contribution considered, our findings indicate strong heterogeneity in internal macroeconomic dynamics related to monetary policy actions as modelled by IR functions. Also, we find evidence for possible existence of similarly behaving (homogenous) subgroups among the economies analyzed. The magnitude and persistent nature of the heterogeneous behaviour found is supported by comparison to previous studies with similar focus (although actual methodology and data used differ), such as Demertzis et. al. [4] or Hušek and Formánek [8].

Our results should not be interpreted as evidence against common European currency from the point of view of OCA theory, as Mundell [10] and many subsequent theoretical studies (such as Desmet [5]) show that international coordination of monetary policies may help minimizing costs as measured by fluctuations in real macroeconomic variables, even in the presence of prominent asymmetric shocks which are closely related to the heterogeneous IR dynamics behaviour identified here. Also, the current euro zone crisis is generally attributed to long term fiscal irresponsibility in some countries, not to monetary policy issues.

Acknowledgements

The research project was supported by Grant No. IGA F4/1/2012, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Blanchard, O. J. and Quah, D.: The dynamic effects of aggregate demand and supply disturbances. *American Economic Review* **79** (1989), 655-673.
[2] Cecchetti, S. G. and Rich, R. W.: *Structural estimates of the U.S. sacrifice ratio*. Federal Reserve Bank of New York, New York, 1999.
[3] Corbo, V.; Landerretche, O. and Schmidt-Hebbel, K. Does Inflation Targeting Make a Difference? In: *Inflation Targeting: Design, Performance, Challenges*, (Loayza, N. and Soto, R., eds.). Central Bank of Chile, Santiago de Chile, 2002, 221–269.
[4] Demertzis, M., Hallet, H. A. and Rummel, O.: Does a Core-Periphery Regime Make Europe into an Optimal Currency Area? In: *European Monetary Union: Transition, International Impact and Policy Options* (Welfens, P. J. J., ed.). Springer-Verlag, Berlin, 1997.
[5] Desmet, K. *Asymmetric Shocks, Risk Sharing and the Latter Mundell*. Banco de Espaňa. Documento de Trabajo No 0222, 2002.
[6] Hušek, R. *Ekonometrická analýza*. Oeconomica, Praha, 2007.
[7] Hušek, R. and Formánek, T.: Estimation of the Czech republic sacrifice ratio for the transition period. *Prague Economic Papers* **1** (2005), 1-15.
[8] Hušek, R. and Formánek, T.: Konvergence české ekonomiky k eurozóně a odhad nákladů přijetí eura v ČR. *Statistika* **1** (2006), 51-63.
[9] Jarociński, M.: *Responses to monetary policy shocks in the East and the West of Europe*. ECB Working Paper No. 970, 2008.
[10] Mundell, R. A. Uncommon Arguments for Common Currencies. In: *The Economics of Common Currencies.* (Johnson, H. G. and Swoboda, A. K., eds.). Allen and Unwin, 1973, 114–132.
[11] 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.

Quality optimization of slab casting with use of software monitoring tool and statistical methods

Zdeněk Franěk [1], Miloš Masařík[2]

Abstract. The paper summarizes the basic analytical and empirical findings obtained during search of dependences of the influence of thermal process taking place during continuous casting of steel semi-finished products – slabs on their quality. The assessment of the quality of slabs uses the collected data and stores them for an effective evaluation of the measured qualitative quantities. The paper describes the proposal of the concept of solution and the function of software for thermal process assessment of the slab casting. Further analyses of possibilities of use of statistical methods for the assessment of the slab casting process are carried on with the possibility of slab quality prediction. The aim of the proposal of the above mentioned process is to provide to technological workers not only the possibility to analyze the process of casting, but also the long-time monitoring and optimization of the production. This approach brings at the process of slab casting important economic benefits.

Keywords: continuous casting, quality assessment, software, statistical methods, regression, optimization, economic benefits.

JEL Classification: C21

AMS Classification: 62J05

1 Introduction to the problem

Quality assessment at the continuous steel casting plant is an inseparable part of the metallurgical plant information system.

This assessment functions on the base of the data necessary for an effective evaluation of the dependence between measured values and quality benchmarks.

The system comprises data recording and filtration, data selection for the matrix of dependent and independent variables, data categorization and statistical method for search of dependences, data presentation and presentation of results of data analyses. The program system enables export of data matrixes into other statistical programs.

Due to the fact that an enormous amount of data is collected during steel casting and that parameters are sampled at the time interval of 10 seconds, the problem of data storage and their effective assessment becomes highly urgent. It is necessary to solve also several other problems in the information system that are related to the quality slab assessment. Firstly, the quality assessment of slabs is the assessment of the semi product, which is determined for further processing, and its insufficient quality appears only after rolling and completion of the final product. Secondly, not all the slabs can be assessed as cold rolled, because slabs are determined to be hot rolled. Therefore the majority of slabs will not pass through the quality inspection. Hence, prediction of the slab quality is needed. In order to predict the slab quality an archive of cast melts with the assessment of slabs is needed, as well as experience of technologists, who are able to create rules for prediction occurrence of possible defects.

For setting of the prediction system it is necessary to create an independent module monitoring the process of casting on the continuous casting machine (CCM), which further sorts out, aggregates and assesses the acquired data in relation to the quality of slabs, and preferably even in the relation to the final product – rolled sheet metal.

2 Motivation

This paper describes a possible approach to creation of such a monitoring program subsystem of the steel plant information system. It is necessary to design a data model with description of data for production monitoring.

[1] Silesian University, School of Business Administration in Karvina, Czech Republic, franek@opf.slu.cz
[2] EVRAZ Vitkovice Steel, Ostrava, Czech Republic, milos.masarik@evrazcz.cz

Basic processing functions assign the production data to the quality benchmarks using the above mentioned data model and the data warehouse. The result is creation of the data warehouse, statistical data analyses and their metallurgical interpretation. Hereunder we describe the monitoring program system LITIOS solving the tasks mentioned above (see figure 1). It has been realized in practice. In the future the data warehouse and the monitoring system [3] will be used for creation of the system of quality prediction.

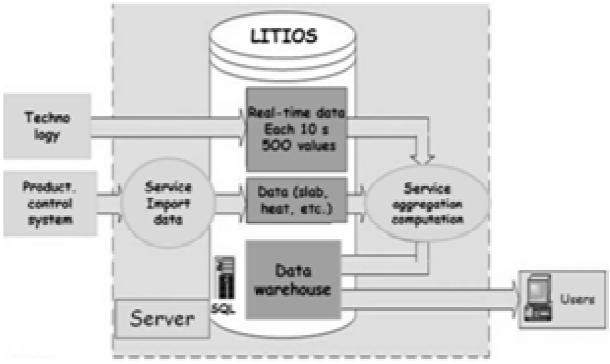


Figure 1: Architecture of the monitoring system LITIOS

3 The goal the project

The monitoring program system LITIOS and creation of the Data Warehouse [1], [4] at continuous casting plant is part of the project "Numerical & Stochastic Model of the Concast Steel Blanks of Rectangular Profile of the Czech Science Foundation", (GA CR Project No. 106/09/0940). The basis of the project consists in obtaining of new findings in solidification thermo kinetics of continuous casting of rectangular steel blanks, and also in creation of a mathematical-statistical method for prediction, numerical optimization and control of the process technology and quality of the final product. An original dynamic model will analyze the temperature field of continuous casting in real time. It will comprise a program for scanning all operational parameters on the caster. Calculation of the temperature field of the slabs will be confronted with the measured parameters on its interface. The stochastic optimization model [2] takes over values from the dynamic model, from the information system of the steelworks and rolling mill, including the data on the quality of the final products. On-line statistics will be ran over these data.

After determining the limit values of the deciding parameters (on the basis of a long-term file of data from the dynamic model) a control system will be elaborated and operation of the caster will be corrected according to the known current status of the caster and recommended limit values. Based on the data from the dynamic model, an on-line comprehensive quality prediction system will be developed, into which only real defects, including the causes of their occurrence, will be integrated. Graphical simulation of the prediction rules will be used to recommend immediate parameters for continuous casting.

4 System design

Since the control hardware and software of the caster are significantly heterogeneous and, furthermore, the casting process is continuous, it is necessary to ensure reliability of the communication tier. The communication tier enters the data in to its own temperature model using the XML/RPC standard with use of the standard TCP/IP protocol. It can easily be adapted to various hardware and software configurations ~~on~~ at the first and second levels of caster control – currently, drivers for Siemens and ABBPLCs exist, as well as for ORACLE and MS SQL databases. Synchronous data (recurring at regular intervals) is recorded every 10 seconds. Other asynchronous information, such as the melt number, chemical analyses, positioning of the tundish and opening of the tundish are read only when a new event occurs. The communication tier checks the validity of the data collected in this way, i.e. it verifies, whether it lies within the acceptable limits, or whether or not it has arrived. Erroneous data is replaced by previous correct data or by the so-called standard values, and then this complete and verified data is passed on to the dynamic on-line model. There are approx. 250 such quantities. The tier of the numerical model carries out the calculation of the temperature field and of other aggregated quantities, which it sends back to the steelworks information system. The temperature model calculates the values of approx. 250 aggregated quantities, so approx. 500 values appear every 10 seconds.

4.1 Data model

For optimizing the production from the point of view of the production quality at the continuous steel casting plant some particular items, so called attributes, were sorted out from the following sets. Based on the selected sets and attributes a diagram of the data model for the program system LITIOS was defined (see Figure 2).

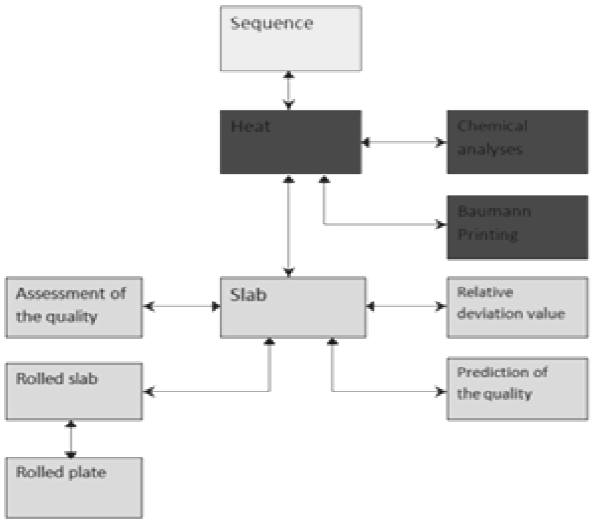


Figure 2 Diagram of the data model

Relational connections were made between the sets. The casting is divided into slabs, and slabs into segments (1m). The set containing the basic data about the slab (selected data from the set "slab production") was chosen as the operative one. For each slab exactly one record is formed. The metallurgical slab code is used as unique key for each record. Relation to another related set is made by the slab code attribute.

4.2 Viewing of functions

The system is modular and it is created on the base of the most advanced knowledge from the field of data technology and methods of data analyses. Views of the data characterising the slab are available for recording and data verification. Functions of the data selection are available in data analyses methods. Data are stored in the matrix of causes (measured production data) and effects (qualitative indicators). Statistical investigation can be made over such reduced data within the frame of the system LITIOS, from which data can be exported into other statistical software, as needed.

Viewing of functions:

- Data visualization and their graphic interpretation
- Searching by selected criteria
- Selection of castings and slabs – setting of the filter
- Selection of displayed items
- Setting of the format of displayed attributes
- Setting of the range for display of deviations of technological values
- Algorithm of search of implicative dependences by the association method
- Verification of hypotheses by the association method
- Exporting of displayed data into the data set (xls, csv, html and alike) for further statistical processing

Description of functions of the program and instructions for manipulation are contained in the user's manual of the system LITIOS.

5 Statistical analysis of dependences between the defects and the technological parameters

Due to the fact that the system LITIOS contains currently a sufficiently large archive set, it is possible to make an investigation of basic dependence by the single linear and quadratic regression, with the prerequisite of the relatively considerable rate of probability finding the dependences. We emphasize that we consider use of the correlation analysis here to be the "introduction into the problem" and further we suppose using more progressive methods of the relation between the parameter analysis and the given defect. Statistical analyses using the

association methods, multiple logistic regression or relatively new branch of neuron nets, which tries to implement the data processing and operating by algorithms adapted from living organisms into current theories, are also available.

It is understandable that for search of relations between the defect and parameters of casting it is possible to use only the slabs, which were processed at the slab finishing mill. It was possible to detect and describe defects, or to state their „non defectiveness". Approximately 19 000 such slabs were available only for the midterm of the last year. That is the set with the large evidence ability even further segmentation can be done. In our case we used four groups steel grades – deep drawing steel, structural steel with different contents of Si, and microalloyed steel.

As there were six types of monitored steels and the number of parameters of casting was twenty three, the total number of possible dependence of the „ steel grade –defect – parameter" was equal to the product of 5x6x23, which represents altogether 690 combinations. It is obvious that testing such a big number of dependences would be technically very demanding and highly time consuming. It therefore was necessary to make selection of dependences in order to reduce their number from the most significant hundreds down to tens of cases.

For making sets with at least basic statistical predicative ability we took the number of cases of detected defects in the group concerned. As the minimum we have chosen the number of 100 defects, thanks to which the number of dependences has decreased several times.

The logical possibility of influencing the occurrence of the specific defect in the particular quality group by the particular parameter of casting was then use of another "sieve ". The dependence of the type „middle rolling crack versus cooling in the mould", „non metallic inclusion versus secondary cooling," etc., were thus excluded from the selection.

5.1 Simple linear or quadratic regression

Before carrying out the analysis of dependence of parameters of casting the input data were also assessed from the point of their objectivity. Descriptive statistics were performed for each parameter based on the minimum and the maximum, or more precisely, on the number of cases in these outer limits. Using this assessment, admissible values of theses parameters were set.

Simple linear and quadratic regression was used for evaluation of dependence of occurrence of the given defect (inception) on the value of the particular casting parameter. From the equation of both of these dependences (linear and quadratic) one was chosen, which expressed the dependence with a greater degree of probability, i.e. the one that had a lower value of the F-test.

In the descriptive statistics the following is always valid:

- R2 - coefficient of reliability. It can oscillate within the limits from 0 to 1 and for the rough check we set that its value must be higher than 0. 5 for being it worth to deal with the dependence at all,
- F - value of the so called F test. In a simplified way it can be said that the value F subtracted from 1 and multiplied by 100 gives the probability indicator of occurrence of the particular dependence. In our case it is ideal, if the value is F = 0.05 or lower, that means that the dependence has the level of probability of at least ~~on~~ 95 % or more.

Values mentioned in the following graphs were obtained always in such a way that the spectrum of occurring values of the particular parameter (independent variable quantities) was into selected intervals (e.g. at the casting speed after 0.1 m.min-1) and in the given interval the percentage of slabs was counted with the found out deficiency. These values were then used as a base for calculation of linear and also parabolic dependence. The equation of dependence with a lower (better) value of the F test is shown in the graph.

5.2 Metallurgical interpretation of some dependences

In this chapter examples of results of the examined dependence are given. By the „metallurgical interpretation" of these results ~~here~~ the authors intend to indicate that even statistical conclusions resulted from big sets must be read and interpreted by an experienced steelworker understanding technological and metallurgical dependences, which cannot always be recorded only by the statistical facts.

The following figures present graph of examples of the examined dependences with the given equation of the particular dependence (linear or quadratic) and the already commented coefficients R2 and F are shown. Graphs are completed by set intervals of input parameters with the percentage of defects in them and with basic values of the descriptive statistics. It is advisable to add the following to these dependences.

In principle oscillation marks are not regarded as a defect in our case. That means the case when the root of the mark is charged by the rolling crack. It is partly difficult and arduous to identify such defects, or as far as our estimation is concerned, there would be just only very few of such cases would be available for statistical analyses. Therefore the criterion (limit) was set up dividing them basically into „smaller and bigger (above limit)". It can be stated in a simplified way that the quality of the slab surface is equal to the depth of the mark, the percentage of slabs in the above limited group can be a very good guidance for setting the optimal range of values of the examined casting parameter.

As for marks it is proper to add in connection with the secondary cooling that the percentage of deficiencies defects shown in the graph does not mean the percentage of the total production burdened by them, but only the percentage of slabs, which showed this crack on the cross section of the slab after its cutting by the flame cutting machine.

Oscillation Marks and the Casting Speed

The quadratic dependence (Tab. 1 and Fig. 3), in contrast to the linear one, shows a great probability measure (F = 0.006), with the minimum around the value of the casting speed 1,3 m.min-1.

Interval from	to	% def.
0.6	0.7	75.0
0.7	0.8	44.4
0.8	0.9	30.3
0.9	1	43.5
1.0	1.1	31.8
1.1	1.2	28.1
1.2	1.3	31.8
1.3	1.4	31.7
1.4	1.5	26.0
1.5	1.6	29.8
1.6	1.7	51.9

Table 1: Percentage of the above limit marks in intervals

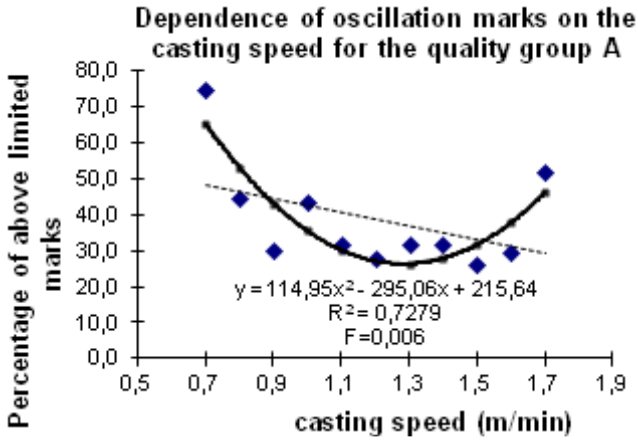


Figure 3: Oscillation marks and casting speed

However, after omission of the limit values (low number of cases), the dependence of the size of oscillation marks on the speed can be interpreted in the way that the size of the marks is in the speed interval from 1.0 to 1.6 m.min-1 nearly constant, favourable, and in the interval from 0.7 to 1.6 m.min-1 it has a slightly declining character (which corresponds also to the dependence of marks on the oscillation). From the point of view of this quality parameter of the slab surface it is therefore proper to avoid casting at the speed below the value of 0.7 m.min-1 (preferably below 1.0 m.min-1) and above 1.6 m.min-1.

Oscillation Marks and Acceleration

The linear dependence (Tab. 2 and Fig. 4) with the great probability measure nearly does not need any commentary.

Interval from	to	% def.
-0.01	0.04	45.4
0.04	0.08	43.8
0.08	0.13	52.9
0.13	0.18	54.1
0.18	0.22	57.6
0.22	0.27	60.3
0.27	0.32	58.6
0.32	0.36	50.0

Table 2: Percentage of oscillations marks in intervals

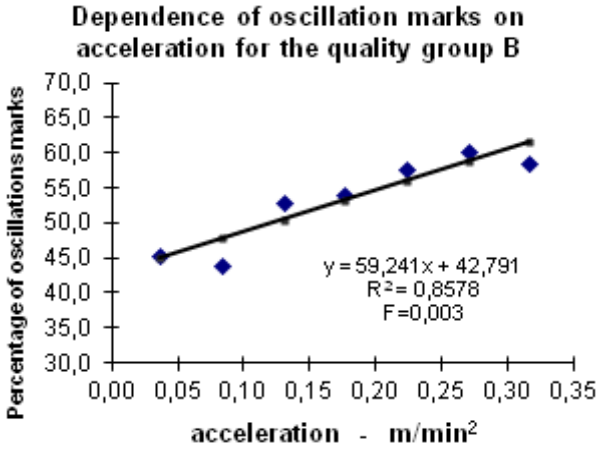


Figure 4: Oscillation marks and acceleration

It totally agrees with the logical reasoning about the influence of acceleration on the quality of the slab surface – increasing value of acceleration decreases (increases) the depth of marks.

Summary of the results

- In the first stage of the solution a total of 16 dependences of defects on the technological casting parameters were searched for. Linear and quadratic regression was used. In 9 cases out of 16 tested relations a statistically significant dependence was found.
- The relation, which seemed to be unreasonable – namely the dependence between formation of middle cracks and oscillation was intentionally investigated. The result was surprisingly "good", that is the biggest "non dependence" from the investigated relations.
- As far as unproven dependence is concerned, it was mainly the dependence of occurrence of non-metallic inclusions on the casting parameters.
- The correlation between the oscillation marks and the casting speed was proved. The dependences determined both for deep drawing steel and also structural steel made it possible to set clearly limits of the casting speed casting and mould oscillation, above which the depth of oscillation marks increases.
- The dependence between the oscillation marks and acceleration is very strong and clear, whereas at the same value of the acceleration a smaller percentage of deficiencies was registered in deep drawing steel.
- The dependence of middle longitudinal cracks on the secondary cooling has a clear and logical trend – with the increasing flow of water (intensity of cooling) the number of cracks on the slabs decreases. At the same time the limit, above which on the contrary a sudden increase of the crack occurrence appears, is clearly marked.

6 Economical benefits of statistical analysis

These results of statistical analysis enable correction of the selected casting parameters, which should reduce (or even eliminate) the number of defects on slabs by selecting the optimal values and limits of these parameters. In practice it brings essential economical benefits. The system of monitoring is useful at continuous production at the steelmaking plant. Long term statistical analysis and control is applied to setting up and updating of optimal values of the casting process parameters.

The above described approach to optimization of production of the process of continuous casting of steel is an inherent part of the information system of the steelmaking plant, and it is supposed to generate the following economical benefits:

- reduction of scrap factor,
- increase of continuous casting production,
- decrease of the production failure risk on the CCM (continuous casting machine),
- possibility of better identification of cracks and of their technological causes,
- necessary pre-requisite for creation of an on-line system for prediction of cracks.

The monitoring system LITIOS, statistical approaches and long term optimization process of production on the CCM form an inevitable part of production optimization both from technological and economical views.

References

[1] Franěk, Z., Kavička, F., Štětina, J., Masarik, M. Software Analytical Instrument for Assessment of the Process of Casting Slabs. In *Proceedings of The 10th International Conference on Numerical Methods in Industrial Forming Processes NUMIFORM 2010*. Pohang, Rep. of Korea 2010. s. 586-592. ISBN 978-0-7354-0800-5.

[2] Mauder, T., Franek, Z., Kavicka, F., Masarik, M. Stetina, J.: A Mathematical & Stochastic Modelling of the Concasting of Steel Slabs, *Conference proceedings 18th International Conference on Metallurgy and Materials*, 2009, 41 – 48, ISBN 978-80-87294-10-9.

[3] Pyszko, R., Cudzik, L., Barabáš, R., Fojtík P., Adamik, M. Advanced Process Monitoring Systems for Continuous Processes. In *Proceedings of the 16th IFAC World Congress*. Prague: Elsevier, June 2006, Paper code: Th-E14-TO/4. 6 p. ISBN-13: 978-0-08-045108-4, ISBN-10: 0-08-045108-X.

[4] Franěk, Z., Masarik, M. Practical Utilization of Analytical Software Instruments at Optimisation of Production of Continuously Cast Slabs. In *Conference proceedings 20th International Conference on Metallurgy and Materials METAL 2011*, Brno [CD]. Ostrava : TANGER, spol. s r.o., 2011. 6 p. ISBN 978-80-87294-22-2.

Incomplete specialization and offshoring across Europe

Richard Frensch [1], Jan Hanousek [2], Evžen Kočenda[3]

Abstract. Recent empirical studies have been searching for evidence on and driving forces for offshoring. We suggest an alternative specification rooted in incomplete specialization that views bilateral gravity equations as statistical relationships constrained on countries' multilateral specialization patterns. Our results support evidence for offshoring activities across Europe, driven by countries' multilateral specialization incentives, as expressed by supply-side country differences relative to the rest of the world.

Keywords: international trade, gravity model, offshoring, panel data, European Union.

JEL Classification: F14, F16, L24
AMS Classification: 62J12, 62P20

1 Introduction

Fragmentation describes the deepening of the division of labor, by horizontally or vertically splitting the production process into distinct tasks. The division of labor encourages specialization and deepening the division of labor thus increases incentives towards specialization, based on either comparative advantage or economies of scale. To realize gains from fragmentation and specialization, it may pay to break up the spatial concentration of production within a firm or even a single plant: firms may outsource tasks. The term offshoring describes the international aspect of this phenomenon, whether or not tasks leave the legal bounds of the firm. Apart from potential gains from specialization, offshoring tasks implies costs of coordinating what is now an international production network rather than a firm or a plant. These coordination or service link costs typically entail costs of investment, communication and of trading inputs to and outputs of offshored tasks, i.e., intermediate products, such as parts and components. It follows that one would expect firms to offshore tasks whenever specialization gains outweigh the implied service link costs, such that the volume of offshoring should increase with fragmentation, or with declining coordination costs, or with the strength of international incentives to specialization.

Specifically, across Europe one would expect the central and east European countries that entered the EU in 2004 as new members (the EU-10) to specialize in labor-intensive tasks and the old EU members (the EU-15) to specialize in capital-intensive tasks, generating two-way trade in intermediate goods across Europe. This process could be expected to be most distinct during and supported by the beginning of the European convergence process. In paper we contribute to the identification of evidence on and driving forces for offshoring activities.

From this description of influences on offshoring, one would expect supply-side country differences to play a role, as in a factor-proportions setting. We theoretically motivate a gravity equation model to analyze gross trade flows related to offshored activities, based on Havemann and Hummels [7]. Our specification is rooted in incomplete specialization, with complete specialization as a natural limiting case, that views bilateral gravity equations as statistical relationships constrained on countries' multilateral specialization patterns. This view reveals countries' multilateral specialization incentives as driving bilateral trade, corresponding to and competing with the role of multilateral trade resistance. Our results support evidence for offshoring activities across Europe, driven by countries' multilateral specialization incentives, as expressed by relative (to the rest of the world) supply-side country differences.

2 Model

When searching for evidence for and determinants of offshoring, a bilateral gravity framework for analyzing gross trade flows related to offshoring activities (i.e., processing trade, trade in parts and components etc.) is set

[1] University of Regensburg and Institut für Ost- und Südosteuropaforschung (IOS) Regensburg. Landshuter Straße 4, 93047 Regensburg, Germany, Department of Economics, frensch@osteuropa-institut.de.
[2] CERGE-EI, Charles University and the Academy of Sciences, Politickych veznu 7, 11121 Prague, Czech Republic; Anglo-American University, Prague; The William Davidson Institute, Michigan; and CEPR, London, jan.hanousek@cerge-ei.cz.
[3] CERGE-EI, Charles University and the Academy of Sciences, Politickych veznu 7, 11121 Prague, Czech Republic; Anglo-American University, Prague; CESifo, Munich; IOS Regensburg; The William Davidson Institute, Michigan; CEPR, London; and the Euro Area Business Cycle Network, evzen.kocenda@cerge-ei.cz.

up in a way that it encompasses an eclectic combination of the determinants spelled out in competing theories to empirically determine, which of them is more important. Apart from exporter and importer market sizes, supply-side country differences are supposed to catch factor proportions influences, similarity measures between countries are to reflect new trade theory or economic geography influences, where similarity measures may even be isomorphic to supply-side country differences. We argue that gravity equations augmented by ad-hoc supply-side country differences are miss-specified since they neglect the key issue of specialization. Factor proportions theories of trade are incomplete specialization models while new theories of trade give way to complete specialization. This difference should result in fundamentally different gravity specifications. According to [7], due to the adding-up constraints of countries' expenditure systems, for trade between more than two countries a combination of four assumptions suffices to derive the simplest possible bilateral gravity structure. These are: (i) trade is only in final goods; (ii) trade is frictionless and balanced; (iii) preferences over final goods are identical and homothetic; (iv) each good is produced in and exported out of only one country, independent from the details on the supply side that give rise to this complete specialization. Under these conditions, bilateral trade is simply a log-linear equation in both countries' incomes, and there is no scope for "augmenting" the gravity equation, e.g. by adding absolute values of differences in per capita incomes.

Maintaining assumption (ii) above, we assume that there are no trade frictions and all trade is balanced. Production is horizontally fragmented in the spirit of Grossman and Rossi-Hansberg [5], where firm-specific production technologies are available to all countries but used by firms in countries rather than by countries. Hence, n tasks are carried out, each of which results in a tradable intermediate good, i.e. a part or a component. One final good is assembled from these n parts or components. Compatible with assumption (iii), all production is subject to homothetic derived demands, such that all variables can be studied in nominal terms: C is consumption or use, X production, Y income, EX exports, and IM imports. Subscripts denote countries, superscripts goods. Given the existence of n intermediate goods and neglecting primary inputs, value-added Z is in each country j distributed over two stages of production:

$$Z_j^k = X_j^k = \delta_j^k Y_j \quad for\ k = 1, \ldots, n \tag{1}$$

and $$Z_j^{n+1} = X_j^{n+1} - \sum_{k=1}^{n} C_j^k = \delta_j^{n+1} Y_j, \ \ with \ \ \sum_{k=1}^{n} \delta_j^k + \delta_j^{n+1} = 1 \tag{2}$$

such that $$\sum_{k=1}^{n} Z_j^k + Z_j^{n+1} = Y_j \tag{3}$$

With homotheticity in production, $$C_j^k = \phi_j^k X_j^{n+1} \ for\ k = 1, \ldots, n \tag{4}$$

With (2) and (3), value added in producing the final good can be written as

$$Z_j^{n+1} = \delta_j^{n+1} Y_j = X_j^{n+1} - \sum_{k=1}^{n} C_j^k X_j^{n+1} - X_j^{n+1} \sum_{k=1}^{n} \phi_j^k$$

$$= X_j^{n+1}\left(1 - \sum_{k=1}^{n} \phi_j^k\right) \tag{5}$$

such that $$X_j^{n+1} = \delta_j^{n+1} Y_j / (1 - \sum_{k=1}^{n} \phi_j^k) \tag{6}$$

$$NE_j^{n+1} = X_j^{n+1} - C_j^{n+1} = \frac{\delta_j^{n+1} Y_j}{1 - \sum_{k=1}^{n} \phi_j^k} - Y_j = \left(\frac{\delta_j^{n+1}}{1 - \sum_{k=1}^{n} \phi_j^k} - 1\right) Y_j \tag{7}$$

For intermediate goods, output is given in (1) and use is in (4), which also holds for the world, $C_w^k = \phi_w^k X_w^{n+1}$. With final goods output as described in (6),

$$\frac{C_j^k}{C_w^k} = \frac{\phi_j^k}{\phi_w^k} \frac{\delta_j^{n+1} Y_j}{\delta_w^{n+1} Y_w} \frac{1 - \sum_{k=1}^{n} \phi_w^k}{1 - \sum_{k=1}^{n} \phi_j^k}, for\ k = 1, \ldots, n$$

This expression can easily be simplified using two characteristics of world trade: first, we know from the world version of (7) that $1 - \sum_{k=1}^{n} \phi_w^k = \delta_w^{n+1}$, as world trade in final goods must be balanced. Second, world output of any good is equal to world use, such that

$$C_j^k = \frac{\phi_j^k}{\phi_w^k} \frac{\delta_j^{n+1}}{(1 - \sum_k \phi_j^k)} \delta_w^k Y_j$$

Country *j*'s net exports of intermediate good k are thus described by

$$NE_j^k = X_j^k - C_j^k, \text{for k} = 1, \ldots, \text{n}$$

Hence, $$NE_j^k = \delta_j^k Y_j - \frac{\phi_j^k}{\phi_w^k}\frac{\delta_j^{n+1}}{1-\sum_k \phi_j^k}\delta_w^k Y_j = \left(\delta_j^k - \frac{\phi_j^k}{\phi_w^k}\frac{\delta_j^{n+1}}{1-\sum_k \phi_j^k}\delta_w^k\right) Y_j \quad (8)$$

As we are only interested in intermediate goods trade, we may simplify (8) by assuming balanced final goods trade for each single country, such that

$$NE_j^k = (\delta_j^k - \frac{\phi_j^k}{\phi_w^k}\delta_w^k) Y_j, \quad \text{for k = 1, …,n} \quad (9)$$

On the basis of (9), countries export an intermediate good if they devote a greater share of value added to producing this good than the rest of the world ($\delta_j^k > \delta_w^k$), or if their intermediate good is more productive in terms of final output than the rest of the world ($\phi_j^k < \phi_w^k$). With firm-specific technologies, identically available everywhere in the world for offshoring activities, as assumed in [5], this simplifies further to,

$$NE_j^k = (\delta_j^k - \delta_w^k) Y_j, \text{ for k = 1, …,n} \quad (10)$$

Summing over all *k*, *j*'s exports of intermediate goods to the world are,

$$NE_j = Y_j \sum_{k=1}^{n} (\delta_j^k - \delta_w^k) \quad (11)$$

Suppose now that intermediate goods are indeed homogeneous. Then, goods are either exported or imported but not both, and positive NE_j indicates a country's exports. Selecting export items with positive net exports into the set K_{EXj}, country *j*'s multilateral intermediate goods exports are,

$$EX_j = Y_j \sum_{k \in K_{EX_j}} (\delta_j^k - \delta_w^k) \quad (12)$$

and are log-linear in income and a specialization pattern, $\sum_{k \in K_{EX_j}} (\delta_j^k - \delta_w^k)$, exhibiting a unitary elasticity with respect to country of origin income, provided the specialization pattern is uncorrelated with income. Analogously for imports,

$$IM_j = Y_j \sum_{k \in K_{IM_j}} (\delta_w^k - \delta_j^k) \quad (13)$$

3 Gravity specification

We argue that countries' bilateral trade under incomplete specialization is driven by multilateral specialization incentives, exactly matching multilateral specialization patterns in form of deviations from world average as described in equations (12) and (13), i.e. in form of countries' deviations of capital-labor ratios (proxied by GDP per capita) or – absent factor price equalization – deviations of wages from the world average.

With incomplete specialization and costless trade, it is not possible to analytically decompose (12) and (13) into bilateral trade relationships. However, trade is not costless, and the way to resolve the indeterminacy is by letting importers choose partners to minimize trade costs. In this sense view bilateral trade equations as statistical relationships constrained on countries' multilateral specialization patterns.

In particular, it is possible to formulate two conditions, subject to which bilateral trade relationships will be distributed in a statistical sense in a sample of countries. First, for bilateral trade to occur, countries' specialization patterns as described in (12) and (13) must be complementary. Second, equations (12) and (13) describe countries' multilateral trade and they can be expected to be met on the average of all bilateral trading relationships.

These two conditions yield predictions for bilateral trade relationships: larger countries trade more in the average of all their trading relationships. Hence, the bilateral trade volume will increase with the product of trading countries' incomes, and countries more specialized vis-à-vis the world can be expected to trade more with each other, provided, their specialization is complementary.

Incentives for incomplete specialization and trade with parts and components are supply-side country differences in factor endowments or wages, wages accurately capture supply-side country differences directly. Consistent with specialization patterns described relative to the world, bilateral trade volumes can be expected to increase with relative supply-side country differences, $|w_j - w_w| \times |w_i - w_w|$, i.e., with the product of countries' respective supply-side differences against the world (w_w). Specifically, within a panel of EU-25 countries, bilateral trade in parts and components ($EX(PC)_{ji,t}$) can be described, without accounting for trade barriers, by the estimable specification that is rooted in our model and that takes the following simple form of a gravity model:

$$log\, EX(PC)_{ji,t} = \beta_0 + \beta_1\, log(Y_{j,t} \times Y_{i,t}) + \beta_2\, log\big(|w_{j,t} - w_{w,t}| \times |w_{i,t} - w_{w,t}|\big) + \\ + \beta_3 Dummy(EU15/10)_{ji}\, log\big(|w_{j,t} - w_{w,t}| \times |w_{i,t} - w_{w,t}|\big) + c_{ji} + k_t + \varepsilon_{ij,t} \quad (14)$$

where *DummyEU15/10* equals one for trade relationships between a EU-15 and a EU-10 country, and zero otherwise.

The specification (14) is estimated on unbalanced panel data with mean time length of about 10 years. In the specification (14) we use time-invariant asymmetric country-pair specific effects (c_{ij}) to capture fixed effects between exporting and importing countries that do not change over time.

The combined variable $DummyEU15/10_{ji}\, log\big(|w_{j,t} - w_{w,t}| \times |w_{i,t} - w_{w,t}|\big)$ is interacted with time-period effects and for this purpose we divide the sample period (1992–2008) into five sub-periods of (almost) equal length. This way we are able to capture technological progress. We then estimate our gravity specification (14) to derive effects on three types of goods: parts and components that represent our primary interest, and two types of final gods, e.g. capital and consumer goods.

We begin our estimation with performing a Hausman-type specification test to assess potential endogeneity of the explanatory variables by comparing a standard fixed effects model with the Arellano and Bover [1] technique. As the test confirms endogeneity of explanatory variables we proceed with instrumentation. We estimate the theoretically motivated specification (14) in a panel setting with fixed effects plus instrument variables a) to overcome problems of omitting variables bias and b) to control for time invariant endogeneity and selection bias.

4 Data

Bilateral trade in parts and components $EX(PC)_{ji}$ describes exports of parts and components from country *j* to country *i* over the period 1992-2008. The data were obtained from the BACI database drawn from the United Nations COMTRADE data as in Frensch and Gaucaite-Wittich [3].

In our estimation we employ three different measures of the bilateral trade in parts and components. First we measure the trade flows of how much country *j* exports to country *i*. Then, following Frensch [2] we measure bilateral trade along the extensive and intensive margins. Trade along an extensive margin, represents *variety* of parts and components of capital goods exported from country *j* to country *i* at time *t*. Trade along the intensive margin, represents *intensity* of parts and components exported from country *j* to country *i* at time *t*. Computations of both extensive and intensive margin measures are performed on the basis of the BACI Database described in Gaulier and Zignago [4].

Further, Y_j and Y_i are exporter and importer GDP at current prices, respectively obtained from the World Development Indicators. Measure of supply-side country differences are wages in exporting (w_j) and importing (w_i) countries and they are measured as annual wage average in manufacturing sector of the exporting (importing) country *j* (*i*) at specific year *t* data were obtained from the International Labor Office statistical databases.

5 Empirical results

We introduce our benchmark results based on specification (14) in the first column of Table 1, where we present estimated coefficients for dependent variables of bilateral parts and components trade. Our results support evidence for offshoring activities generating trade in parts and components of capital goods due to the existence of multinational production networks across Europe, and inform about driving forces identified already in the first section.[4]

Statistically significant coefficients β_1 demonstrate that larger countries trade more with each other. Second, negative coefficients β_2 confirm that our sample of European countries on average in fact features a rather homogeneous specialization pattern as compared to the world average. However, comparing coefficients β_2 and β_3 points to relative supply-side country differences as driving offshoring activities across Europe compatible with models of incomplete specialization and trade, specifically between original EU-15 and the ten accession countries (EU-10), rather than within each of the two country groups. Third, technical progress in terms of declining service link costs and ongoing fragmentation – as captured by the sub-period dummies – appears to positively

[4] Importance of the international trade between the EU10 and EU15 countries have been documented via its effect on stock markets. Specifically, macroeconomic announcements on the EU current account affect prices on the Czech, Hungarian, and Polish stock markets even at intra-day frequency (see Hanousek and Kočenda [6] for details).

influence offshoring: with the exception of the final sub-period, for EU-15/EU-10 pairs, β_3 is increasing slowly over time. The slight decrease of the β_3 coefficients in the final 2005–2008 sub-period might indicate that EU-10 countries catch up with the EU-15 so that supply-side country differences between both groups, relative to the world, become less pronounced. This may well be affected by the technological progress in the EU-10 countries that is closely linked to foreign direct investment and multinationals (Uzagalieva et. al. [8]). As foreign-owned subsidiaries become a part of the innovation systems and industrial structure of the EU-10 countries they promote overall technological growth in the region that further contributes to the catch-up with the EU-15.

		Flows	**Extensive Margin**	**Intensive Margin**
log Y_j Y_i		0.718***	0.254***	0.464***
		(0.023)	(0.013)	(0.014)
log ($\|w_j - w_w\| \times \|w_i - w_w\|$)		-0.101***	-0.040***	-0.061***
		(0.020)	(0.010)	(0.013)
log ($\|w_j - w_w\| \times \|w_i - w_w\|$) for EU-15 / EU-10 pairs	1992-1995	0.183***	0.104***	0.079***
		(0.036)	(0.020)	(0.021)
	1996-1998	0.202***	0.117***	0.085***
		(0.036)	(0.019)	(0.021)
	1999-2001	0.241***	0.145***	0.096***
		(0.035)	(0.019)	(0.020)
	2002-2004	0.251***	0.157***	0.094***
		(0.034)	(0.018)	(0.020)
	2005-2008	0.230***	0.132***	0.099***
		(0.033)	(0.018)	(0.020)
N		27,354	27,354	27,354

Table 1 Parts and components, *w*=wages (simple world averages)

Finally, we have performed several robustness checks. Our results are robust with respect to a different measure of country differences (GDP per capita) as well as population weighted averages. Further, our results reveal reveals that trade in parts and components due to offshoring activities across Europe is predominantly realized along the intensive margin in response to market size increases, but along the extensive margin in response to stronger relative supply-side country differences. I.e., more offshoring of activities from the EU-15 to the EU-10 in response to stronger relative supply-side country differences means predominantly offshoring of new activities rather than extending the scale of already offshored activities.

6 Conclusions

We argue that analyzing gross trade flows related to offshored activities by using gravity equations augmented by *ad hoc* measures of supply-side country differences appear miss-specified. We develop gravity framework, rooted in incomplete specialization that views bilateral gravity equations as statistical relationships constrained on countries' multilateral specialization patterns, allowing for offshoring to increase with fragmentation and declining coordination costs, with multilateral incentives to specialization, and to decline with multilateral trade resistance.

We apply this framework to a truly Europe-wide sample of countries, while fully accounting for potential tendencies towards factor price equalization *via* trade, and find evidence for offshoring activities across Europe driven by countries' multilateral specialization incentives, as expressed by relative (to the rest of the world) supply-side country differences. In particular, the results do not contradict those provided by Grossman and Rossi-Hansberg [5], and are thus compatible with the view that offshoring need not hurt (low-skill) workers, as long as offshoring relationships get strengthened along the intensive margin as opposed to the extensive margin by new relationships. Our results, however, suggest that exactly this latter might have been the case recently when extending offshoring from the EU-15 to the EU-10.

Acknowledgements

We benefited from comments by Gordon Hanson and Jürgen Jerger, and presentation participants at Regensburg, FU Berlin, WHU Vallendar, wiiw Vienna, Perugia, UC San Diego, and WU Wien. We thank Anastasiya Shamshur and Stephan Huber for their excellent research assistance. Jan Hanousek and Evžen Kočenda acknowledge research support from the GAČR grant No. 403/12/0080. The usual disclaimer applies.

References

[1] Arellano, M., and Bover, O.: Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* **68** (1995), 29–51.
[2] Frensch, R.: Trade liberalisation and import margins. *Emerging Markets Finance & Trade*, **46** (2010), 4–22.
[3] Frensch, R., and Gaucaite-Wittich, V.: Product variety and technical change. *Journal of Development Economics* **88** (2009), 242–257.
[4] Gaulier, G., and Zignago, S.: BACI: International Trade Database at the Product-level. The 1994-2007 version. *CEPII working paper* **23** (2010).
[5] Grossman, G. M., and Rossi-Hansberg, E.: Trading tasks: a simple theory of offshoring. *American Economic Review* **98** (2008), 1978–1997.
[6] Hanousek, J., and Kočenda, E.: Foreign News and Spillovers in Emerging European Stock Markets. *Review of International Economics* **19** (2011), 170–188.
[7] Havemann, J., and Hummels, D.: Alternative hypotheses and the volume of trade: the gravity equation and the extent of specialization. *Canadian Journal of Economics* **37** (2004), 199–218.
[8] Uzagalieva, A., Kočenda, E., and Menezes, A.: Technological Innovation in New European Union Markets. Forthcoming in *Emerging Markets Finance and Trade*, (2012).

Quantitative evaluation of life quality of Czech districts

Ludvík Friebel[1], Jana Friebelová[1]

Abstract. We have evaluated the quality of life of the LAU1 regions in the Czech Republic, according to the selected criteria by the Data Envelopment Analysis (DEA) method. This method was initially proposed to evaluate the efficiency. In this paper, the level of efficiency represents the level of life quality. The efficiency (quality of life) is in this case represented as a share of output in weighted sum of inputs. In other words, it represents a certain degree to which desirable output can offset undesirable indicators. Initially, we considered four types of input: unemployment rate, criminality, average length of incapacity to work, index of ageing and one output: proportion of economic active inhabitants. In the next stage, the fifth input, the average price of a dwelling, is added. Its impact on changes of the efficiency score is under consideration. Reached results were then compared with net migration.

Keywords: data envelopment analysis, district, LAU1, unemployment, criminality, incapacity of work, index of ageing, dwelling, migration.

JEL Classification: C02, C44, C60
AMS Classification: 90C05, 90B50

1 Introduction

Many authors have dealt with quality of life assessment recently. Life quality at the level of administrative districts (LAU1) of South Moravia is investigated, for example, by Živělová and Jánský [11]. In their work, life quality is assessed on the base of analysis of the population and unemployment increase. Moreover, the authors involve indicators of medical care and transport and technical infrastructure.

The assessment of regions resp. regional disparities is considered by the Ministry of Regional Development of the Czech Republic within the scope of the WD-05-07-3 – Regional disparities Program in the availability and affordability of housing, their socioeconomic consequences and tools directed to decrease of regional disparities, for details see Lux and Sunega [7].

To analyse life quality in particular districts in our work, we 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 districts 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 is considered to the greatest extent, see more in section Material and Methods.

In the paper by authors Martic and Savic, [9], the assessment of the regional performance in Serbia was conducted using the DEA Models together with the 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 [10] 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 [6] using the DEA Method.

We evaluated the efficiency of 14 districts (LAU 1) in the Czech Republic, according to the selected criteria by the DEA Method. For this contribution, we selected district of NUTS2 - Southwest. This method was initially proposed to evaluate the efficiency. In the first stage, we considered four types of input: unemployment rate, criminality, average length of incapacity to work, index of ageing and one output: proportion of economic active inhabitants. In the next stage, the fifth input, the average price of a dwelling, is added. Its impact on changes of the efficiency score is under consideration.

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, [5]) 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

[1] University of South Bohemia, Faculty of Economics, Department of Applied mathematics and informatics, Studentská 13, 370 05 České Budějovice, Czech Republic, friebelova@ef.jcu.cz, friebel@ef.jcu.cz.

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 described in detail, for example, in work by 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 Southwest NUTS2. 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 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

$$\sum_{j=1}^{n} y_{jH} u_{jH} - \sum_{i \in ND} v_{iH} x_{iH} \tag{1}$$

subject to

$$\begin{aligned} &\sum_{j=1}^{n} y_{jk} u_{jH} \le \sum_{i \in ND} x_{ik} v_{iH} + \sum_{i \in D} x_{ik} v_{iH}, \; k = 1,2,..., p, \\ &\sum_{i \in D} x_{iH} v_{iH} = 1, \\ &u_{jH} \ge \varepsilon, \\ &v_{iH} \ge \varepsilon (i \in D), \\ &v_{iH} \ge 0 (i \in ND). \end{aligned} \tag{2}$$

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 dependent 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). A non-efficient 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 this reason, we involved in efficiency assessment only four resp. five inputs and one output.

In our work, we dealt with life quality on the level of districts (LAU 1). Considering the data accessibility, we assessed regions from year 2008 to year 2010. We tried to involve the economic and social factors into assessment and the reached results were then compared with net migration.

Malmquist index is used to assessment of the development of efficiency in time. It was formulated, in its first form, already in 1953 by Swedish economist Sten Malmquist [8]. The index enables to distribute the efficiency development into two parts, change of efficiency of a unit with respect to other assessed units and change of production possibilities frontier caused by so called technological progress.

In the first stage, 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 economic active inhabitants (proportion of inhabitants aged from 15 to 64 to the whole population). The source data are available at http://www.czso.cz.

In the next stage, we added another controllable input – the average price of dwelling in the given districts. This entry was obtained from http://www.realitymorava.cz, where the comparison of dwelling prices in all districts of the Czech Republic is periodically published. We monitored how this input influences efficiency and contribution changes of particular indicators. More closely, we observed, to what extent the low price of dwelling in ineffective districts can compensate weak points of the other inputs.

By the means of the model with the above mentioned inputs and output, we derived efficiency of particular districts.

In the last stage, we compared efficiency of districts with net migration within three considered years. We supposed that the efficiency of districts would correspond closely with net migration.

3 Results

The source data are available at http://home.ef.jcu.cz/~friebel/research/districts/ dea_2.xls. Data were processed in program Maple using our own application. Functionality of the application was verified by SW Frontier Analyst. Maelmquist indices in case of non-discretionary CCR model can be computed with DEAP. For this purpose, we created our own application in Maple.

In Table 1, we can see efficiency obtained for three assessed years. For evaluating the efficiency progress. we calculate also Malmquist index for particular districts. Remote sparsely populated districts of West Bohemian region e.g. Klatovy, Tachov are less efficient. On the contrary, the central districts of this region are fully efficient through the assessed period. In South Bohemian region is the situation different, where there are inefficient populous districts e.g. Tábor or Písek.

	Efficiency			Malmquist	
	2008	**2009**	**2010**	**2008/09**	**2009/10**
Č. Budějovice	1.000	1.000	1.000	1.000	1.000
Č. Krumlov	1.000	1.000	1.000	1.000	1.000
J. Hradec	1.000	1.000	1.000	1.000	1.023
Písek	0.973	1.000	0.960	1.014	1.006
Prachatice	1.000	1.000	1.000	0.964	1.000
Strakonice	0.863	0.881	0.797	1.011	0.969
Tábor	0.954	0.989	0.922	1.018	1.004
Domažlice	1.000	1.000	1.000	1.000	1.021
Klatovy	0.925	1.000	0.916	1.039	0.988
Plzeň - město	1.000	1.000	1.000	1.000	1.000
Plzeň - jih	1.000	1.000	1.000	1.000	1.000
Plzeň - sever	1.000	1.000	1.000	1.000	1.000
Rokycany	0.903	0.903	0.920	0.959	1.107
Tachov	0.880	0.909	1.000	1.016	1.049
Average	0.964	0.977	0.965	1.001	1.012

Table 1 Efficiency scores and Malmquist indices w/o price of dwelling

In Table 2, we can see efficiency obtained for three assessed years including prices of dwelling. The mentioned price is given as an average value in the selected year in a particular district. This inclusion shifts the district of Tachov to the efficiency frontier. That trend is evident from Malmquist index.

When including dwelling into the assessment, it helps the district of Strakonice to reach the efficiency frontier, but only for year 2008. The higher efficiency of districts, the more balanced Malmquist indices.

	Efficiency			Malmquist	
	2008	**2009**	**2010**	**2008/09**	**2009/10**
Č. Budějovice	1.000	1.000	1.000	1.000	1.000
Č. Krumlov	1.000	1.000	1.000	1.000	1.000
J. Hradec	1.000	1.000	1.000	1.000	1.011
Písek	0.973	1.000	0.960	1.014	1.006
Prachatice	1.000	1.000	1.000	1.000	1.000
Strakonice	1.000	0.952	0.797	0.976	0.920
Tábor	1.000	0.989	0.922	0.995	1.004
Domažlice	1.000	1.000	1.000	1.000	1.007
Klatovy	0.925	1.000	0.916	1.039	0.988
Plzeň - město	1.000	1.000	1.000	1.000	1.000
Plzeň - jih	1.000	1.000	1.000	1.000	1.000
Plzeň - sever	1.000	1.000	1.000	1.000	1.000
Rokycany	0.967	0.906	0.925	0.968	1.099
Tachov	1.000	1.000	1.000	1.000	1.000
Average	0.990	0.989	0.966	0.999	1.002

Table 2 Efficiency scores and Malmquist indices icluding price of dwelling

In Table 3, we can see efficiency obtained for three assessed years including average prices of dwelling given as an average price for the all districts in selected year. This computation was made in order to consider influence of addition next input into assessment. This experiment is discussed in detail in the concluding part of the contribution.

	Efficiency			Malmquist	
	2008	**2009**	**2010**	**2008/09**	**2009/10**
Č. Budějovice	1.000	1.000	1.000	1.000	1.000
Č. Krumlov	1.000	1.000	1.000	1.000	1.000
J. Hradec	1.000	1.000	1.000	1.000	1.023
Písek	0.979	1.000	0.977	1.011	1.014
Prachatice	1.000	1.000	1.000	1.000	1.000
Strakonice	0.980	0.981	0.968	1.001	1.030
Tábor	0.984	0.991	0.978	1.003	1.031
Domažlice	1.000	1.000	1.000	1.000	1.021
Klatovy	0.982	1.000	0.979	1.009	1.022
Plzeň - město	1.000	1.000	1.000	1.000	1.000
Plzeň - jih	1.000	1.000	1.000	1.000	1.000
Plzeň - sever	1.000	1.000	1.000	1.000	1.000
Rokycany	0.992	0.985	0.991	0.997	1.038
Tachov	1.000	1.000	1.000	1.000	1.000
Average	0.994	0.997	0.992	1.001	1.013

Table 3 Efficiency scores and Malmquist indices including average price of dwelling

Comparison between efficiency and migration is depicted in Table 4. Because of elimination of short-term fluctuation of migration we considered average values in term 2008-2010. It is interesting that districts with negative net migration are fully efficient.

	Efficiency 2008-2010	Net migration 2008-2010
Č. Budějovice	1.000	843.333
Č. Krumlov	1.000	-33.000
J. Hradec	1.000	140.000
Písek	0.978	162.333
Prachatice	1.000	-77.667
Strakonice	0.847	35.000
Tábor	0.955	115.667
Domažlice	1.000	237.667
Klatovy	0.947	128.000
Plzeň - město	1.000	1135.000
Plzeň - jih	1.000	600.667
Plzeň - sever	1.000	702.667
Rokycany	0.909	220.667
Tachov	0.930	106.000
Average	1.000	843.333

Table 4 Relation between efficiency and net migration (average values within 2008 and 2010)

4 Conclusion

All 14 districts were assessed according to the following criteria: unemployment rate , criminality, average length of incapacity to work, index of ageing and one output proportion of economic active inhabitants. Seven districts in year 2008 were inefficient. Next year, one of them reached full efficiency, but in the following year (2010), it fell back into the group of inefficient districts. The average efficiency of the whole group was 0.964 in year 2008; 0.977 in year 2009; and 0.965 in the last considered year.

In the next stage, we added another input – the average price of dwelling in the given districts. We monitored how this input influences DMU's efficiency. Our assumption was that low price of dwelling in some regions increase their efficiency. On the contrary, overpriced flats in some districts can decrease efficiency. This precondition was met in the case of all previously inefficient districts. In year 2008, one of them turned to full efficient, the rest of them increased efficiency. In the following years, only three regions remained inefficient with higher efficiency (the average efficiency in that year was 0.989). The changes in efficiency in the last assessed year was marginal compared to the assessment excluding price of dwelling. The average efficiency in year 2010 was 0.966, only one district increased efficiency

Including price of dwelling into the set of criteria, the average efficiency naturally rises. That fact can be produced by including additional input into assessment. In order to exclude this factor, we made another assessment with the average prices of dwelling for all considered districts. The average efficiency in this case is higher the previous one. For year 2008, we obtained 0.994; for 2009 – 0.997 and for year 2010 – 0.992.

Namely, the district of Rokycany in the year 2008 increased efficiency to 0.967 after including dwelling into inputs, but with including the average price, which is lower than the real price, its efficiency is higher than the previous one - 0.992. The district of Strakonice in the year 2009 increased efficiency in the same way to 0.952, but with lower average price its efficiency increased to 0.981.

It means that including real price of dwelling does not increase efficiency of regions as much as including the average prices of dwelling. Therefore, we can say that price of dwelling could not compensate weakness of districts in other indicators.

Finally, we tried to determine a mutual connection between these criteria and net migration. Based on the previously obtained results, we considered efficiency without prices of dwelling. Because of elimination of short-term fluctuation of migration, we considered the average values in the period of 2008-2010. We can say that districts with lower average efficiency are less attractive than districts that are fully efficient. However, the greatest migration was found at the central districts, for example České Budějovice, Plzeň etc. On the contrary, the highest loss of inhabitants is from inefficient regions, i.e. Tachov, Prachatice, Strakonice, Klatovy. An exception represents the district of Český Krumlov.

We have to remember that people making decision where to live also consider other aspects which are not included in this presented assessment. These aspects can be objective, measurable criteria, but also the characteristics and nature of these people. For sure, their decisions can be also influenced by the place where their fami-

lies live, where they were studying etc. People, who have already settled down somewhere, such as managers or physicians, are willing to move.

Acknowledgements

This work was supported by Ministry of Education, Youth, and Sport of Czech Republic.

References

[1] Andersen, P., Petersen, N. C.: A procedure for ranking efficient units in DEA. *Management Science*, Vol. 39 (1993), 1261–1264.
[2] Banker, R. D., Charles, A., Cooper, W. W.: Some models for estimating technical and scale inefficiencies in DEA. *Management Science*, Vol. 30, No. 9 (1984), 1078–1092.
[3] Charnes, A., Cooper, W. W., Rhodes, E.: Measuring the efficiency of decision making units. *European Journal of Operational Research*, Vol. 2, No. 6 (1978), 429–444.
[4] Cooper, W. W., Seiford L. M., Tone K.: *Data Envelopment Analysis. A Comprehensive Text with Models, Applications, References and DEA-Solver Software.* Springer, New York, 2007.
[5] Farrell, M. J.: The measurement of productive efficiency. *Journal of the Royal Statistical Society*, Vol. 120, No. 3 (1957), 253–290.
[6] Li, ZG., Cheng, Y.: 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, 2010, 1315–1319.
[7] Lux, M., Sunega, P.: Vliv podmínek bydlení na zamýšlenou migraci české populace za prací. *Czech Sociological Review – Czech,* Vol. 43, No. 2 (2007), 305–332.
[8] Malmquist, S.: Index numbers and indifference surfaces. *Trabajos de Estadistica*, 4, (1953), 209–242.
[9] Martic, M, Savic, G.: 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 (2001), 343–356.
[10] Xiong, GQ., Liu, YQ., Tang, BQ.: 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, 2008, 12748–12751.
[11] Živělová, I., Jánský, J.: Analysis of life quality development in the administrative districts of South Moravia. *Agric. Econ. – Czech*, Vol. 54, No. 9 (2008), 431–439.

Polygon regular location problem

Lýdia Gábrišová[1], Jaroslav Janáček[2]

Abstract. The polygon regular location problem originates in public transportation system management, where a coordination of circulating vehicles must be performed so that the vehicle arrivals are distributed regularly in some period. The contribution deals with the problem complexity and reports on various ways of reducing the computational effort, which is necessary to obtain an optimal solution of the problem using a general optimization environment.

Keywords: regular polygon, location problem, public transport, coordination of vehicle arrivals, free order of objects.

JEL Classification: C61
AMS Classification: 90C27

1 Introduction

The problem originated in the field of public transport as we can see in [2, 4, 8]. An original goal was to increase attractiveness of public transport by making schedule of urban and sub-urban transport more regular at some selected stops. It was taken into account that a regularity of vehicle arrivals optimizes a transportation supply for passengers by non-investment way, which can be seen in [1, 3, 5, 8]. It was found that individual vehicles as buses, trams or trolleybuses circle along their lines in the associated urban transportation network and in addition an average time of traversing a cycle is relatively short. Under these circumstances, the same vehicle usually appears at an observed stop several times in a given period. All the vehicle arrivals form a transportation supply for the passengers coming at the stop. If some arrivals follow closely one after other, the second one of the arrivals does not contribute considerably to the transportation supply.

On the other hand, long intervals between arrivals cause an unpleasant time loss for passengers, which come at the stop randomly. It follows that some regularity of the arrivals is desirable. The regularity of vehicle arrivals at the given stop can be improved by a shift of arrival time of an individual vehicle. As a given vehicle appears at the stop several times in the given period, a shift of one of its arrivals causes shifts of all its arrivals in the period by the same value. Furthermore, it must be considered that the observed stop can be served by vehicles, which traverse different lines. It causes that time intervals between neighboring arrivals of different vehicles differ. If the observed period is long enough to be divisible by circle time of each considered vehicle, then arrival times of a given vehicle can be depicted as vertices of a regular polygon on a circle, whose circumference is equal to the length of the period. That is for; we can call the problem as location of vertices of polygons on a circle or briefly the regular polygon location problem. In the problem, the goal is to locate the set of regular polygons in a circle so that all vertices lie on the same circumference and their distribution be regular [6, 7, 11].

Many researchers tackled this problem in several recent decades but due to non-linearity and discreteness of associated models, only heuristics have been used to solve this problem. In this contribution, we present an exact approach based on usage of particular characteristics of the problem and thorough model building. This approach together with new possibilities offered by used optimization environment enable us to solve some instances of the problem to optimality, which can be seen in [9, 10, 13].

2 Reduced Formulation of the Regular Polygon Location Problem

Let us consider r regular polygons with the same radius and center. It follows that all polygon vertices lie on the circumference. Let the p-th polygon have n_p vertices. Vertex locations of the polygon p on the circle are uniquely given by an angle between a zero point on the circle and the first vertex of the polygon. Let T denote the circumference of the circle given in some angle units and let $d_p=T/n_p$ hold. If we introduce a decision variable x_p, which denotes the angle between the zero point and the first vertex of the p-th polygon, then the second vertex has location d_p+x_p, the third vertex has location $2d_p+x_p$ and so on.

[1] University of Žilina, Faculty of Management Science and Informatics, Department of Mathematical Methods, Univerzitná 2, 010 26 Žilina, Slovak Republic, email: lydia.gabrisova@fri.uniza.sk
[2] University of Žilina, Faculty of Management Science and Informatics, Department of Transportation Networks, Univerzitná 2, 010 26 Žilina, Slovak Republic, email: jaroslav.janacek@fri.uniza.sk.

In general the variable x_p corresponds with a value of rotation of the p-th polygon from a zero point. We also define a reverse mapping $p(k)$, which returns the index of polygon containing the vertex k (see figure 1). The total number of involved vertices is denoted as m. It is obvious that range $< 1, d_p >$ is sufficient for x_p, $p \geq 2$, to cover all possible locations of the p-th polygon vertices taking into consideration the condition that no vertex location is allowed to meet other vertex locations.

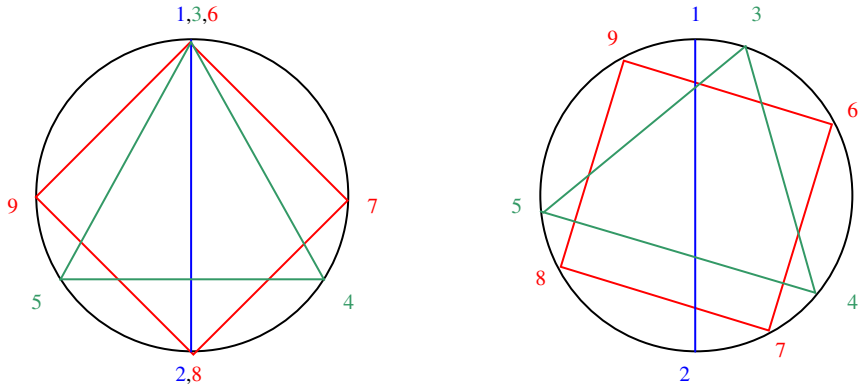


Figure 1 Example of three polygons $n_1 = 2$, $n_2 = 3$, $n_3 = 4$ vertices with designed labelling of vertices. Left figure – the default state and right figure – possible rotation of polygons, where the first polygon is fixed.

Now we can assign the lowest value of rotation a_k to each vertex $k = 1, \ldots, m$ and state that current value of rotation of vertex k is given by $a_k + x_{p(k)}$ and this location varies over range $< a_k, a_k + d_{p(k)} - 1)$. Without loss of generality, we can set the values of a_1 and x_1 at zero. The regular polygon location problem can be formulated as a search for such vector $< 0, x_2, \ldots, x_r >$, which corresponds to the most regular distribution of the vertices along the period T (circumference). The regularity is considered as a sum of squares of the differences between neighboring vertex locations in this contribution. If we denote t_k as circumference distance between the vertex k and the directly preceding vertex, then sum of $(t_k)^2$ for $k = 1, \ldots, m$ corresponds with convex nonlinear objective function. The k-th item of the objective function can be linearized in accordance to [12].

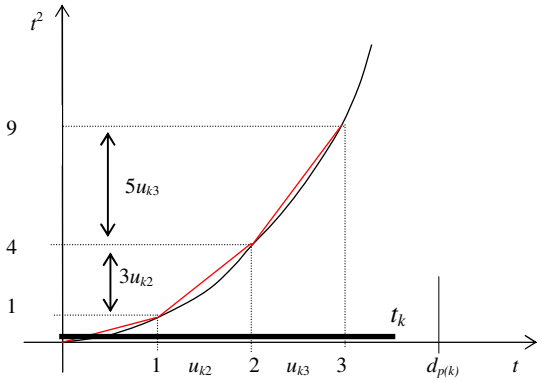


Figure 2 Piecewise linearization of the item $(t_k)^2$.

For t_k, $k = 1, \ldots, m$ we introduce non-negative auxiliary variables $u_{kj} \leq 1$ for $j = 2, \ldots, d_{p(k)}$. Then we can express t_k as the sum of u_{kj} in accordance to (1) as depicted in figure 2. The square of t_k can be replaced by linear expression depending on variables u_{kj} accordingly to (2).

$$t_k = 1 + \sum_{j=2}^{d_{p(k)}} u_{kj} \quad for\ k=1,..,m \tag{1}$$

$$(t_k)^2 = 1 + \sum_{j=2}^{d_{p(k)}} (2j-1)u_{kj} \qquad for\ k=1,..,m \tag{2}$$

If the values of u_{kj} are integer, then the right-hand-side of (2) represents the exact value of $(t_k)^2$. Let index of the vertex preceding the vertex k be denoted as $i(k)$. We realize that the variable x_1 was set to zero. This way, t_1 can be defined as $T - a_{i(1)} - x_{p(i(1))}$, where $i(1)$ is index of the vertex with the biggest value of location in the

period T. The other variables t_k must satisfy the link-up constraints $t_k = a_k + x_{p(k)} - a_{i(k)} - x_{p(i(k))}$ for $k = 2, \ldots, m$. Unfortunately the order of vertices changes by jumps, when the values of x_p vary.

The non-linearity involved in precedence mapping (permutation) $i(k)$ used in the substitution constraints for t_k can be removed by introducing auxiliary zero-one variables $w_{ik} \in \{0, 1\}$ for each relevant pair (i,k), $i = 1, \ldots, m$, $k = 1, \ldots, m$, $i \neq k$. A variable w_{ik} takes the value of one if and only if the vertex i directly precedes the vertex k. To describe the relevant pairs in the following model, we introduce a logical function *exists* defined on all pairs $(i,k) \in \{1, \ldots, m\} \times \{1, \ldots, m\}$. The function $exists(i,k)$ takes the value of true, if and only if the pair (i,k) is relevant. Making use of the previously introduced variables x_p and u_{kj} a reduced linear model of the regular polygon location problem can be formulated as follows:

$$\textit{Minimize} \qquad \sum_{k=1}^{m} (1 + \sum_{j=2}^{d_{p(k)}} (2j-1) u_{kj}) \tag{3}$$

$$\textit{Subject to} \qquad \sum_{\substack{i=1 \\ exists(i,k)}}^{m} w_{ik} = 1 \quad for\ k=1,..,m \tag{4}$$

$$\sum_{\substack{i=1 \\ exists(k,i)}}^{m} w_{ki} = 1 \quad for\ k=1,..,m \tag{5}$$

$$x_1 = 0,\ x_p \leq d_p - 1 \qquad for\ p=2,..,r \tag{6}$$

$$T - a_i - x_{p(i)} \geq 1 + \sum_{j=2}^{d_{p(1)}} u_{1j} - T^L_{i1} * (1 - w_{i1}) \quad for\ i=2,..,m,\ exists(i,1) \tag{7}$$

$$T - a_i - x_{p(i)} \leq 1 + \sum_{j=2}^{d_{p(1)}} u_{1j} + T^U_{i1} * (1 - w_{i1}) \quad for\ i=2,..,m,\ exists(i,1) \tag{8}$$

$$a_k + x_{p(k)} - a_i - x_{p(i)} \geq 1 + \sum_{j=2}^{d_{p(k)}} u_{kj} + T^L_{ik} * (1 - w_{ik})\ for\ k=2,..,m,\ i=1,..,m,\ exists(i,k) \tag{9}$$

$$a_k + x_{p(k)} - a_i - x_{p(i)} \leq 1 + \sum_{j=2}^{d_{p(k)}} u_{kj} + T^U_{ik} * (1 - w_{ik})\ for\ k=2,..,m,\ i=1,..,m,\ exists(i,k) \tag{10}$$

$$u_{kj} \leq 1,\ u_{kj} \geq 0 \qquad for\ k=1, ..., m,\ j=2,..,d_{p(k)} \tag{11}$$

$$x_p \in Z^+ \qquad for\ p=1,..,r \tag{12}$$

$$w_{ik} \in \{0,1\} \quad for\ i=1,..,m,\ k=1,..,m,\ exists(i,k) \tag{13}$$

The consistency constraints (4) and (5) ensure that each vertex k has its predecessor and successor. The constraints (7) – (10) cause that if $w_{ik} = 1$ for some pair (i, k), then the difference between the location of vertex k and the location of preceding vertex i is equal to t_k given by substituting equality (1). If $w_{ik}=0$ holds, then the associated constraints are relaxed by suitable values of T^L_{ik} and T^U_{ik} .

3 Model Adjustments

In the model above, there are several loose ends, which must be set up before submitting an associated instance to a general IP-solver for solving. The first thing, which must be determined is the logical function *exists(i,k)*. The cardinality of the set of pairs *(i,k)*, for which the function takes the value of "*true*" corresponds with the number of binary variables w_{ik} and this number forms a considerable part of the instance size. As a general IP-solver performs common branch-and-bound method, it is obvious that complexity of the problem puts a tight limit on size of solved problem instances. The set of variables can be defined in several ways.

The basic approach introduces a variable w_{ik} for each pair *(i,k)*, where $i \neq k$. The approach with reduced model introduces a variable w_{ik} for each pair *(i,k)*, where inequality $a_i \leq a_k + d_{p(k)} - 1$ holds. The advanced approach defines a variable w_{ik} for each pair *(i,k)*, where the constraints from the reduced approach is satisfied and, in addition, there exists no vertex j for which the following inequalities $a_i + d_{p(i)} \leq a_j$ and $a_j + d_{p(j)} \leq a_k$ hold.

The second loose end represents the lower and upper bounds of T^L_{ik} and T^U_{ik} respectively. These coefficients may influence the starting lower bound in the branch-and-bound computational process as the associated lower bounding uses *LP*-relaxation of the model.

The third loose end is the input order (numbering) of the polygons, especially the polygon, whose x_p can be fixed to zero. We can order the polygons either in descending or ascending order accordingly to their number n_p of vertices. In the first case, the polygon with the biggest number of vertices is fixed and in the second case the polygon with the smallest number can be fixed.

At the end of this section, we note that if we do not insist on the condition that no vertex location can share locations of the other vertices, then the constraints (7) and (9) can be relaxed. The original model will be called the full model and the reduced one will be referred as the half model. In the next section we try to show, how these possible settings may influence computational time of the used IP-solver.

4 Numerical Experiments

The presented numerical experiments are aimed at inspecting a special phenomenon, which occurred, when preliminary experiments [12] were performed. It was found that a reduction of decision variables did not impact the computational time proportionally. The numerical experiments were performed with two pools of instances, where each of the pools contains exactly six items. Each instance consists of one quadruple of polygons defined on a circle with circumference T=360. The total number of vertices included in one instance varies from 14 to 20 and from 21 to 27 for the first and second pools respectively. All experiments were performed using the optimization software FICO Xpress 7.1 (64-bit, release 2010). The associated code was run on a PC equipped with the Intel Core i5 2430M processor with the parameters: 2.4 GHz and 4 GB RAM.

The first series of experiments was performed with the full model. The values of T^L_{ik} and T^U_{ik} were set at the biggest value and descending order of polygons was used. The tested models differ only in the number of w_{ik}, where three cases "basic", "reduced" and "advanced" are distinguished in accordance to the denotation introduced in the previous section. Each problem instance was solved to optimality and the associated average computation times and average numbers of introduced variables w_{ik} are reported in table 1.

		Basic		**Reduced**		**Advanced**	
Pool	**Range**	***Avg_CT***	***Avg_w_ij***	***Avg_CT***	***Avg_w_ij***	***Avg_CT***	***Avg_w_ij***
1	14-20	7.7	282	10.4	171	13.2	144
2	21-27	79.6	571	14.6	330	26.5	217

Table 1 Average computational times in seconds and average numbers of introduced variables are reported in columns denoted as Avg_w_{ij} and *Avg_CT* respectively.

The next portion of experiments was focused on the influence of the less or more tight adjustment of the bounds T^L_{ik} and T^U_{ik} on the computational time. The associated experiments were performed with the full model; the number of variables w_{ik} was reduced in accordance to the approach "advanced" and where descending order of polygons was used. Accordingly to the definition in the previous section, the values of the T^L_{ik} and T^U_{ik} were subsequently set at the compromise setting. Each instance of the pools was solved to optimality and the associated average computational times are plotted in table 2.

			Biggest s.	**Tight s.**	**Compromise s.**
Pool	**Range**	***Avg_w_ij***	***Avg_CT***	***Avg_CT***	***Avg_CT***
1	14-20	144	13.2	9.9	2.0
2	21-27	217	26.5	27.0	3.3

Table 2 Average computational times in seconds and average numbers of introduced variables are reported in columns denoted as Avg_w_{ij} and *Avg_CT* respectively.

The last portion of experiments concerns an impact of the possible reduction of constraint set and the polygon order to the performance of branch-and-bound search. In these experiments the biggest setting of T^L_{ik} and T^U_{ik} was applied and the approach "advanced" to introduction of variables w_{ik} was used. The ascending order was tested on the full model and the half model was combined with descending ordering of polygons. The resulting average computational times are reported in table 3.

Model			Full	Full	Half
Ordering			Descending	Ascending	Descending
Pool	Range	Avg_w_{ij}	*Avg_CT*	*Avg_CT*	*Avg_CT*
1	14-20	144	13.2	3.4	374.2
2	21-27	217	26.5	7.5	2261.4 *)

Table 3 Average computational times in seconds and average numbers of introduced variables are reported in columns denoted as Avg_w_{ij} and *Avg_CT* respectively. *) The average was computed from five instances only. Computation of the sixth instance exceeded one hour and was prematurely terminated.

5 Conclusions

To our great surprise the obtained numerical results show that neither variable reduction nor constraint relaxation in model of the polygon regular location problem accelerate common branch-and-bound search embedded in an optimization environment. The only exception is the reduction from "basic" to "reduced" for the pool with bigger number of vertices. The similar results appear in the attempts for tighter setting of relaxing constants T^L_{ik} and T^U_{ik}. Nevertheless, it was found that a convenient setting, which compromises the loose and tight settings, might considerably reduce the computational time of the searching process. Furthermore, the ascending order of polygons in the input data also influences a computational time reduction.

Acknowledgements

This work was supported by the research grant of the Faculty of Management Science and Informatics, University of Žilina VEGA 1/0296/12 "Public Service Systems with Fair Access to Service" and APVV-0760-11 "Designing of Fair Service Systems on Transportation Networks". We would like to thank to "Centre of excellence in computer sciences and knowledge systems (ITMS 26220120007) for built up infrastructure, which was used.

References

[1] Brucker, P. and Hurink, J.: A railway-scheduling problem. In: *Zeitschr. f. Oper. Res.*30 (1986), A 223-A 227.

[2] Burkard, R. E.: Optimal schedules for periodically recusrring events. In: *Discrete Applied Mathematics* 15 (1986), 167-180.

[3] Černá, A.: Optimization of periodic and transport offered. *Scientific Papers of the University of Pardubice, series* B 4 (1998), 193-194.

[4] Černý, J.: Optimization of Periodic Time Tables in Railway and Urban Transport. In: *Proceedings of "Optimization in Production and Transportation", Scheveningen* (1994)

[5] Černý, J. and Blaško, R.: A Note to Coordination of Periodic Processes in Transport. In: *TRANSCOM* 95, *University of Žilina* (1995), 95-98.

[6] Černý, J. and Guldan, F.: Location of polygon vertices on circles and its application in transport studies. *Apl. mat.* 32 (1987), 81-95.

[7] Černý, J. and Kluvánek, P.: *Basises of mathematical transport theory.* Veda, Bratislava, 1991 (in Slovak).

[8] Černý, J. and Mojžíš, V.: Note on the coordination of periodic processes in transportation systems. *Communications–Scientific Letters of the University of Žilina* 7/2 (2005), 25-27.

[9] Gábrišová, L. and Janáček, J.: The solution of some problems with free order of objects with IP-Solver. In: *Proceedings of workshop "Problems of Discrete Optimization for Transport", Univerzity of Pardubice* (2012), (to apear).

[10] Gábrišová, L. and Kozel, P.: Accuracy of linear approximation used in non-linear problem optimization. *Journal of Information, Control and Management Systems of the University of Žilina Vol* 8, No 4 (2010), 301-309.

[11] Guldan, F.: Maximization of distances of regular polygons on a circle. *Apl.Mat.*25 (1980), 182-195.

[12] Janáček, J. and Gábrišová, L.: Regular polygon location problem with IP- Solver. In: *Proceedings of conference Quantitative methods in economic, University of Economics, Bratislava* (2012), 103-107.

[13] Kozel, P. and Gábrišová, L.: Minimalization of passengers waiting time by vehicle schedule adjustment. In: *TRANSCOM* 2011*, University of Žilina* (2011), 103-106.

A causal relationship between foreign direct investment, economic growth and export for Central and Eastern Europe

Zuzana Gallová[1]

Abstract. Foreign direct investment is generally considered to be an instrument how to stimulate economic growth of any country. For this purpose governments of transition countries try to encourage the inflow of foreign direct investment by various measures. The aim of this paper is to analyse the relation between foreign direct investment, economic growth and export in eight countries of Central and Eastern Europe. Estimation of effects on economic growth and export was performed for each country of the region individually in the period from 1993 to 2010. The co-integration method and vector error correction model were applied on quarterly data. The results confirm the existence of long-term causal links between variables studied in five of the eight countries of the region. The impact of foreign direct investment within the region of Central and Eastern Europe, however, is not clear, since there were positive as well as negative effects proven on export.

Keywords: foreign direct investment, export, economic growth, Central and Eastern Europe, co-integration.

JEL Classification: F41, F43
AMS Classification: 62P20

1 Introduction

The paper is aimed to analyse long-term causal relations between export, economic growth and foreign direct investment (FDI). It is assumed that there might be a long-term link between these variables. Export is considered to be the determinant for economic growth. Export and growth of economy openness might lead to the growth of output level and increase of economy growth. As stated in theory, it is the foreign direct investment that contributes to the export performance increase of a country. Such effect happens when we speak about the export oriented FDI. In addition, positive impact on the economy growth of a country is attributed to the foreign direct investment according to the theory. These facts are pointed out by, for example Dritsaki et al. [2], Feridun and Sissoko [4], Pacheco-Lopéz [5].

The paper is divided into six chapters. First chapter is the introduction. The second chapter is aimed at the relevant bibliography overview. A model used and data are specified in the third chapter. The fourth chapter deals with the long-term links model between foreign direct investment, export and gross domestic product. The fifth chapter is about the vector error correction model. The last chapter includes causal relation model results between selected variables.

2 Bibliography overview

There is a series of empirical studies examining FDI effects on economic growth and export and relations between these variables. Such effects are examined by various approaches. The results of individual studies vary, which depends on the period selected, data processed, other variables included in the model or it depends on the econometrics. In respect to this, there is the examination of one way relationship or two way causal relationships. The result of such activity might be to find the one way, two-way or lack of causality. VAR autoregressive model, regression analysis, as well as panel data analysis are used in order to examine relationships among the given variables. The results of selected relevant studies are included in this sub-chapter.

VAR autoregressive model was used to examine the relationships between FDI, export and economic growth in the empirical study by Dritsaki et al. [2]. The research was performed for Greece by means of annual data during 1960 – 2002. The results of study point out the two-way relationship between export and economic growth. Moreover, the impact of foreign direct investment on export, as well as on economic growth in Greece was proven. Fabry [3] examines the relationship between foreign direct investment, export and economic growth by means of Johansen test for cointegration and Granger causality test. The research was performed on a sample of countries from Central and Eastern Europe. According to the research, the impact of foreign direct investment

[1] Silesian University in Opava, School of Business Administration in Karvina, Department of Finance, Univerzitni nam. 134/3, 733 40 Karviná, Czech Republic, gallova@opf.slu.cz.

on economic growth was proved in Albania and Russia. On the contrary, the impact of economic growth on FDI inflow was proved in case of Hungary, Poland and Romania. The author states at the end of the research that export has stronger impact on the economic growth than it has on foreign direct investment in Central and Eastern Europe and, on the contrary, the impact of foreign direct investment on export has not been proved by the research in countries of Central and Eastern Europe. Pacheco-Lopéz [5] who tested annual data from 1970 – 2000 between foreign direct investment and export in Mexico took similar approach. Based on the VAR model she proved the two-way relationship between export and foreign direct investment. On the one hand, export stimulates the foreign direct investment; on the other hand the foreign direct investment contributes to the country's growth of export. Pelinescu and Radulescu [6] deal with impact of foreign direct investment on economic growth in Romania. They use data of gross domestic product, foreign direct investment and export, which are quarterly calculated by logarithm and they are season adjusted, for the period 2001 – 2009. In order to model causal relations they use the same approach as already mentioned in the above researches. In the study they conclude that FDI have slight, however positive influence on both the gross domestic product and export. Furthermore they state that in order to have stronger positive FDI impact on economic growth and export it is necessary to use longer time interval.

3 Model specification and data

Dritsaki et al. [2] use the above mentioned study from the methodology point of view of the causal relations model between exports, gross domestic product and foreign direct investment. Causal relation between the variables shall be examined by VAR autoregressive model in the following form:

$$EXP = f(FDI, GDP) \tag{1}$$

Individual variables in an equation are: EXP = export, FDI = foreign direct investment, GDP = gross domestic product.

Export represents the export of goods and services at regular prices. Foreign direct investment defines the FDI condition in a country. Gross domestic product is expressed in market prices. Each data is in million EUR. These are quarterly data for different time periods depending on the data availability in each country. Time period in Poland is for Q1. 2004 – Q4. 2010. Time period in the Czech Republic and Latvia is for Q1. 2000 – Q4. 2010. In Estonia it is for the time period of Q1.1996 – Q4. 2010. Time period selected for Lithuania, Slovakia and Slovenia is for Q1. 2001 – Q4. 2010. The research in Hungary was performed for a period of Q1. 1995 – Q4. 2010.

The data describing export and gross domestic product are obtained from Eurostat. The data on FDI condition are gained from the individual countries' central banks. Export and gross domestic product data are season clear. Season clear series is marked with "sa" at the end of the time series' title. Individual data were calculated by logarithm before the testing. Calculation by logarithm was performed for the purpose of the time series smaller dispersion and consequently to ensure stationarity of the time series. Individual time series calculated by logarithm are marked with capital letter "L" before the each time series' title.

3.1 Unit root test

In order to perform cointegration that shall be used to test the long-term causal relations between selected variables it is necessary for the logarithmized time series being stationary on the first difference I(1) and nonstationary on its own values. Stationarity test is performed by Augmented Dickey – Fuller test (ADF test). Lag length of the time series in the ADF test was based on the Schwarz criterion. According to the development of logarithm adjusted data, a test stationary equation included a coefficient in case of FDI and EXP and it included a trend coefficient in case of gross domestic product. This is demonstrated by the following equation:

$$\Delta X_t = \delta_0 + \delta_1 t + \delta_2 X_{t-1} + \sum_{i=1}^{k} \alpha_i \Delta X_{t-i} + u_t \tag{2}$$

ADF test is used to determine a unit root X_t on the level of each variable calculated by logarithm in time *t*. Variable ΔX_{t-i} determines the first difference with lag length and u_t suggests the autocorrelation of the error. Coefficients δ_0, δ_1, δ_2 a α_i are suggested. Null and alternative hypothesis for the existence of a unit rook in variable X_t is: H_0: $\delta_2 = 0$, H_ε: $\delta_2 < 0$ (Dickey and Fuller [1]).

Results are shown in the Table 1. On the left side you can find the data determining time series which are not stationary in level value. On the right side there are data determining stationarity of time series with the first difference. On the level of importance, an index "a" means 1% importance, "b" means 5% importance and "c" means 10% importance. The assumption for further test and research of long term relationships between specified variables is met since the time series stationarity was proved in the first differences.

Country	Variable	In their levels		1st differences	
		Lagged	T-statistics ADF	Lagged	T-statistics ADF
CZE	LEXP_sa	9	(-1.194)	9	(-4.904)[a]
	FDI	9	(-2.095)	9	(-4.706)[a]
	LGDP_sa	9	(-1.192)	9	(-4.457)[a]
EST	LEXP_sa	10	(-1.670)	10	(-4.968)[a]
	LFDI	10	(-0.713)	10	(-7.121)[a]
	LGDP_sa	10	(-1.133)	10	(-2.884)[b]
HUN	LEXP_sa	10	(-1.854)	10	(-4.864)[a]
	LFDI	10	(-1.950)	10	(-8.674)[a]
	LGDP_sa	10	(-0.589)	10	(-6.398)[a]
LVA	LEXP_sa	9	(-0.497)	9	(-3.966)[a]
	LFDI	9	(-0.872)	9	(-4.055)[a]
	LGDP_sa	9	(-1.159)	9	(-3.086)[b]
LTU	LEXP_sa	9	(0.701)	9	(-4.788)[a]
	LFDI	9	(-1.257)	9	(-4.707)[a]
	LGDP_sa	9	(-1.430)	9	(-4.174)[a]
POL	LEXP_sa	6	(-1.045)	6	(-2.843)[b]
	LFDI	6	(-1.830)	6	(-4.120)[a]
	LGDP_sa	6	(-1.675)	6	(-3.715)[b]
SLO	LEXP_sa	9	(-1.299)	9	(-3.251)[b]
	LFDI	9	(-2.188)	9	(-5.781)[a]
	LGDP_sa	9	(-0.100)	9	(-4.795)[a]
SVK	LEXP_sa	9	(-0.985)	9	(-3.720)[a]
	LFDI	9	(-2.861)	9	(-6.748)[a]
	LGDP_sa	9	(-0.655)	9	(-4.046)[b]

Table 1 ADF – Unit Root Test

4 Long term relationship test between FDI, EXP and GDP

Johansen test for cointegration was used to test long term relationships between FDI, EXP and GDP. It is necessary to define appropriate time lag length within this test. Here, an Akaike criterion was used while determining the appropriate lag length, which was applied for the non-differentiated VAR model estimation. Two periods with an appropriate lag length were proved in Hungary, Latvia, Poland, Slovenia and Slovakia. Four periods with appropriate lag length were proved in Estonia. Three periods were proved in the Czech Republic and one period in Lithuania.

Long term relationships test between FDI, EXP and GDP in individual countries was performed on the basis of the following equation (3):

$$LEXP\_sa = \alpha + \beta_1 LFDI + \beta_2 LGDP\_sa + \mu \tag{3}$$

The dependent variable is export and the independent variables are FDI and GDP. Cointegration relationships were examined for each country individually. Long term relationships between variables in Johansen test are examined on the basis of two tests, and that is a Trace test and Max-eigenvalue test. The results are shown in

the Table 2. The Table 3 includes cointegration equations. There was no long term relationship between the variables proved according to the Johansen cointegration test in the Czech Republic, Lithuania and Latvia.

Country	Null Hypothesis	Trace Statistic	Critical Value 0.05	Max-Eigen Statistic	Critical Value 0.05
CZE	r=0	38.25707	35.19275	24.66499	22.29962
	r<=1	13.59208	20.26184	9.150793	15.8921
	r<=2	4.441288	9.164546	4.441288	9.164546
EST	r=0	35.33130	35.19275	14.40979	22.29962
	r<=1	20.92152	20.26184	12.87337	15.89210
	r<=2	8.048144	9.164546	8.048144	9.164546
HUN	r=0	63.44122	35.19275	38.12878	22.29962
	r<=1	25.31244	20.26184	19.68283	15.89210
	r<=2	5.629603	9.164546	5.629603	9.164546
LVA	r=0	38.37530	35.19275	23.25614	22.29962
	r<=1	15.11916	20.26184	11.76007	15.89210
	r<=2	3.359081	9.164546	3.359081	9.164546
LTU	r=0	43.42952	35.19275	32.48876	22.29962
	r<=1	10.94076	20.26184	8.710273	15.89210
	r<=2	2.230486	9.164546	2.230486	9.164546
POL	r=0	41.07743	35.19275	24.74526	22.29962
	r<=1	16.33217	20.26184	10.99543	15.89210
	r<=2	5.336741	9.164546	5.336741	9.164546
SLO	r=0	43.09568	35.19275	19.59360	22.29962
	r<=1	23.50207	20.26184	15.35180	15.89210
	r<=2	8.150273	9.164546	8.150273	9.164546
SVK	r=0	38.65633	35.19275	22.70358	22.29962
	r<=1	15.95274	20.26184	11.13284	15.89210
	r<=2	4.819906	9.164546	4.819906	9.164546

Table 2 Johansen cointegration test Variables LEXP, LFDI and LGDP

Country	Cointegration equation
CZE	No cointegration relationship
EST	*LEXP_SA=-0.490LFDI+2.576LGDP_sa-8.835* *(0.480) (0.958) (3.709)*
HUN	*LEXP_sa=3.208LFDI-3.309LGDP_sa+7.786* *(0.529) (0.929) (3.975)*
LVA	No cointegration relationship
LTU	No cointegration relationship
POL	*LEXP_sa=0.067LFDI+1.178LGDP_sa-3.751* *(0.117) (0.199) (0.932)*
SLO	*LEXP_sa=0.840LFDI-0.588LGDP_sa+6.211* *(0.318) (0.858) (4.887)*
SVK	*LEXP_sa=-0.102LFDI+1.367LGDP_sa-2.800* *(0.095) (0.138) (0.431)*

Table 3 Cointegration Equation

5 Vector error correction model

Long term relationships between variables were proved in five countries out of eight; however, cointegration does not take into account short term deviations. For this reason there is a Vector error correction model (VECM) used to detect such deviations within the cointegration.

Vector error correction model was used in such situations where the existence of cointegration relationships was proved. On the basis of the test, an appropriate lag length was defined for two periods (Hungary, Poland, Slovenia and Slovakia), four periods (Estonia). A suitable model setting was tested by selected autocorrelation, normality and heteroscedasticity test. The test proved that neither of the effect was present in the model, it means that the model is set correctly. The results of vector error correction are shown in the Table 4.

Variable	**HUN**	**POL**	**SLO**	**SVK**	**Variable**	**EST**
	D(L_EXP_sa)	**D(L_EXP_sa)**	**D(L_EXP_sa)**	**D(L_EXP_sa)**		**D(L_EXP_sa)**
CointEq1	-0.022	0.399	0.050	0.033	CointEq1	-0.347 [a]
	(-1.180)	(1.087)	(0.594)	(0.189)		(-2.481)
D(L_EXP_sa(-1))	0.476	0.643	0.772	0.431	D(L_EXP_sa(-1))	0.303
	(3.109)	(1.433)	(2.728)	(2.216)		(1.924)
D(L_EXP_sa(-2))	0.061	-0.284	0.101	0.292	D(L_EXP_sa(-2))	0.311
	(0.396)	(-0.642)	(0.383)	(1.469)		(2.117)
D(L_FDI-1))	-0.030	0.039	-0.109	-0.103	D(L_EXP_sa(-3))	0.156
	(-0.398)	(0.113)	(-0.964)	(-0.720)		(1.007)
D(L_FDI-2))	0.061	0.033	0.037	0.305	D(L_EXP_sa(-4))	-0.125
	(0.775)	(0.095)	(0.314)	(2.168)		(-0.813)
D(L_GDP_sa(-1))	0.080	0.017	-0.600	0.138	D(L_FDI-1))	0.157
	(0.448)	(0.027)	(-0.925)	(0.342)		(1.665)
D(L_GDP_sa(-2))	-0.330	-0.089	-0.549	-1.094	D(L_FDI(-2))	-0.096
	(-1.985)	(-0.183)	(-0.920)	(-2.730)		(-0.981)
C	0.015	0.017	0.019	0.029	D(L_FDI(-3))	0.021
	(2.295)	(1.167)	(1.876)	(1.284)		(0.223)
R-squared	0.284	0.515	0.379	0.471	D(L_FDI(-4))	0.093
Adj. R-squared	0.190	0.315	0.229	0.339		(0.981)
Sum sq. resids	0.074	0.029	0.030	0.045	D(L_GDP_sa(-1))	0.502
S.E. equation	0.037	0.041	0.032	0.040		(1.832)
F-statistic	3.012	2.580	2.534	3.571	D(L_GDP_sa(-2))	0.201
						(0.734)
					D(L_GDP_sa(-3))	-0.704
						(-2.711)
					D(L_GDP_sa(-4))	0.144
						(0.493)
					C	-0.002
						(-0.265)
					R-squared	0.522
					Adj. R-squared	0.371
					Sum sq. resids	0.089
					S.E. equation	0.046
					F-statistic	3.455

Table 4 Vector Error Correction Model

The Table 4 shows t-statistics in the brackets. Index “a” means the coefficient importance of the error correction (CointEq1) on the 1% importance level. From the statistic point of view it is proved that this coefficient is

important in equations with the explained EXP variable only in case of one country – Estonia. In other countries such importance of the model was not proved. According to this we can state that, in respect to short term deviations, the model fails to explain adequately the convergence for long term balance in countries where the vector error correction coefficient is not important from the statistic point of view. Consequently, the next chapter shall include the results of causal relationships model between FDI, EXP and GDP only for those countries where the coefficient proved to be statistically important.

6 Results of the causal relationships model between FDI, EXP and GDP

On the basis of the above test, the existence of long term relationships between variables was proved in five countries. Subsequently, the results of VECM proved statistical importance of the correction model for the dependent variable only in Estonia. Arising from such findings, the following sub-chapter includes the analysed results of the causal relationships model for Estonia.

On the basis of cointegration equation for Estonia, as follows:

$$LEXP_{sa} = \underset{(0.480)}{-0.490}LFDI + \underset{(0.958)}{2.576}LGDP_{sa} - \underset{(3.709)}{8.835} \tag{4}$$

the long term positive relationship in the country was not proved between foreign direct investment and export. Based on the equation, as a consequence of FDI growth by 1 % with the lag length of four months the export decreased by 0.49%. It means that foreign direct investment in Estonia do not contribute to the export growth. This might be due to the FDI types that are aimed to seek markets. This FDI types flow into the country with the aim to get a part on a market abroad and reduce the costs to supply such market. Such FDI are not export oriented and for this reason they do not contribute to the export growth.

The statistical importance of the error model coefficient in the chapter five proves that the model can explain the short term dynamics, as well as convergence for the balance condition. In case of Estonia and EXP dependent, the results of adjusted coefficient are high and they prove that 34.7% of short term deviations from the balance condition are adjusted by changes in the model dependent variable with the lag length of four quarters. The result is that the rate of convergence towards the balance condition is very satisfying in this case.

On the basis of the research method and by means of available time series, the generally accepted opinion about the FDI positive effect on foreign trade of a country was not proved. The method of research led to the result analysis of causal relationship models between variables only within one country – Estonia. Other countries failed to prove statistic importance of the VECM correction model and consequently the results analysis of causal relationships model between FDI, EXP and GDP was not performed.

Acknowledgements

Research behind this paper was supported by the Student Grant Competition of Silesian University within the project SGS 25/2010 'Financial integration in the EU and its effect on corporate sector'.

References

[1] Dickey, D.A., and Fuller, W.A.: Distributions of the estimators for Autoregressive time series with a unit root. *Journal of the American Statistical Association* **366** (1979), 427 – 431.

[2] Dritsaki, M., Dritsaki, C., and Adamopoulos, A.: A Causual Relationship between Trade, Foreign Direct Investment and Economic Growth for Greece. *American Journal of Applied Sciences* **3** (2004), 230–235.

[3] Fabry, N. H.: The role of inward – FDI in the transition countries of Europe. In: *Enterprise in transition.* Faculty of Economics Split, Split, 2001, 1032-1055.

[4] Feridun, M., and Sissoko, Y.: Impact of FDI on Economic Development: A Causality Analysis for Singapore, 1976 – 2002. *International Journal of Economic Sciences and Applied Research* **1** (2011), 7-17.

[5] Pacheco-Lopéz, P.: Foreign Direct Investment, Exports and Imports in Mexico. In: *Proceedings of the University of Kent Studies in Economics 0404.* University of Kent, Kent, 2004, 1-24.

[6] Pelinescu, E., and Radulescu, M.: The Impact of Foreign Direct Investment on the Economic Growth and Countries´Export Potential. *Journal for Economic Forecasting* **4** (2009), 153-169.

Simulation-assisted Horvitz-Thompson statistic and isotonic regression

Wojciech Gamrot[1]

Abstract. Under some finite-population sampling schemes the calculation o exact inclusion probabilities may be prohibitively complex even for modest population sizes. This is especially true for various sequential procedures used in spatial sampling and for fixed-cost (or sum-quota) schemes. Such a phenomenon presents a significant challenge for constructing estimates of finite population totals based on the Horvitz-Thompson approach. Such a challenge may be overcome by replacing unknown first-order inclusion probabilities with estimates computed in a simulation study which is enabled by the knowledge of the sampling scheme. Such estimates may be calculated in several ways, which influences stochastic properties of the Horvitz-Thompson statistic. Available auxiliary information may also be used to improve their accuracy. In this paper isotonic regression algorithms are applied to capitalize on limited auxiliary information and to improve the accuracy of simulation-assisted design-based estimates for finite population totals.

Keywords: empirical inclusion probability, Horvitz-Thompson estimator, simulation, population total

JEL classification: C83
AMS classification: 62D05

1 Introduction: empirical Horvitz-Thompson estimation

Consider a finite population represented by a set of indices $U = \{1, ..., N\}$. Values $y_1, ..., y_N$ of a fixed characteristic correspond to each population unit. The parameter under study is the population total:

$$t = \sum_{i \in U} y_i \tag{1}$$

In order to estimate it, an unordered sample s is drawn from U through some sampling scheme characterized by inclusion probabilities of the first-order: $p_1, ..., p_N$ where $p_i = Pr\{i \in s\}$ for $i \in U$. If inclusion probabilities are known then the population total may be estimated without design bias using the well-known Horvitz-Thompson (H-T) statistic [9]:

$$\widehat{t} = \sum_{i \in s} \frac{y_i}{p_i} \tag{2}$$

However, for some sampling schemes exact calculation of first-order inclusion probabilities $p_1, ..., p_N$ may be impossible because of prohibitive computational complexity. This is particularly true for various sequential sampling schemes like those described in [2] and [4] where the combinatorial explosion prevents the computation of inclusion probabilities even for very modest population sizes. At first sight, this effect seems to make the H-T estimation impossible. Fortunately, as noted in [5],[15], another potent source of information still remains in the hands of a statistician. Namely, the sampling procedure itself. This may be used to generate a large number of independent sample replications. By examining sample membership indicator values corresponding to any particular population unit within all replications one may arrive at the estimate of associated inclusion probability. This estimate may take the form of sample proportion of ones or some other function of membership indicators. Let R be the number of replications and let f_i be a number of times a certain i-th population unit is drawn to a sample replication. A classic approach

[1]University of Economics in Katowice, 1 Maja 50, 40-287 Katowice, Poland, e-mail: wojciech.gamrot@ue.katowice.pl

proposed in [6] would rely on estimating the inclusion probability p_i through the statistic:

$$\hat{p}_{iF} = \frac{f_i + 1}{R + 1} \tag{3}$$

for $i \in s$ and inserting these estimates instead of p_i into (2). However, as noted in [6], the number of sample replications needed to guarantee a desired level of accuracy for population total estimates may still be uncomfortably large, even with respect to contemporary computing capabilities. One possibility of overcoming such a difficulty is the adoption of sequential methods at the simulation stage as proposed in [6]. In this paper another approach to empirical H-T estimation based on the use of external information is investigated.

Let us assume that available auxiliary information takes form of an ordering constraint on first-order inclusion probabilities so that these probabilities are known to behave monotonically. Such a situation arises for many sampling schemes such as Pareto sampling discussed in [13] where inclusion probability grows with increasing values of some known auxiliary variate. It also appears in the case of fixed-cost sum-quota sampling proposed in [11], where inclusion probabilities decrease with growing cost of observing the variable. It is worth noting, that when subsequent replications are generated, the values of sample membership indicators may be recorded for all population units, as opposed to observing only units in the sample s. The additional computational effort associated with recording all of them is negligible. Then various isotonic regression algorithms may be applied to enhance the accuracy of inclusion probability estimates within the sample by forcing the ordering constraints to be satisfied by all the estimates in the population. Hence such approach may be viewed as a special indirect case of "strength-borrowing" technique discussed in [10]. In the next section one such algorithm is presented.

2 The PAVA procedure

The well-known Pool-Adjacent-Violators Algorithm (PAVA) works in the following way (see [1], [8], [3]). Let $p_1, p_2, ..., p_N$ be unknown probabilities satisfying a simple order:

$$p_1 \leq p_2 \leq ... \leq p_N \tag{4}$$

Let R_i independent trials be made of an event with probability p_i and let f_i denote the number of successes in these trials ($i = 1, ..., N$). Constraint-preserving estimates $\hat{p}_1, ..., \hat{p}_N$ of $p_1, ..., p_N$ satisfying (1) are calculated by iteratively grouping (merging) initial unconstrained estimates $f_1/R_1, ..., f_N/R_N$ (sample proportions of ones) into groups and repeatedly averaging them within each group. The procedure works through following steps .

Step one: assign the index of each probability p_i for $i = 1, ..., N$ to a separate group so that initial groups are $A_1^{(0)} = \{1\}, ..., A_N^{(0)} = \{N\}$ and initial number of groups is $a^{(0)} = N$. Set an initial estimate of mean probability in each i-th group to $q_i^{(0)} = f_i/R_i$ for $i = 1, ..., N$.

Step two: in each subsequent step of the procedure (numbered $m = 1, 2, ...$) whenever mean probability estimates in some neighboring groups are found to breach the order constraint, a maximum-length sequence $A_i^{(m-1)}, ..., A_{i+z}^{(m-1)}$ (where $1 \leq i < i + z \leq a^{(m-1)}$) of such groups is merged together so that

$$A_i^{(m)} = A_i^{(m-1)} \cup ... \cup A_{i+z}^{(m-1)}$$

and

$$A_j^{(m)} = A_{j+z}^{(m-1)}$$

for $j = i + 1, ..., a^{(m)}$ while $a^{(m)} = a^{(m-1)} - z$. Then a new within-group mean probability estimate:

$$q_i^{(m)} = \frac{\sum_{j \in A_i^{(m)}} f_j}{\sum_{j \in A_i^{(m)}} R_j}$$

is assigned to the group $A_i^{(m)}$ while $q_j^{(m)} = q_{j+z}^{(m-1)}$ for $j = i + 1, ..., a^{(m)}$. This step is repeated until all estimates $q_1^{(m)}, ..., q_{a^{(m)}}^{(m)}$ of mean within-group probabilities satisfy ordering constraints or there is just one group left.

Step three: when the iteration stops after the last - say M-th - step ($M \in \{1, 2, ...\}$) a mean probability estimate computed for a group is assigned to each of its member components so that the final estimate for the probability p_i is $\hat{p}_i = q_j^{(M)}$ for $i \in A_j$, $j = 1, ..., a^{(M)}$.

If $f_1, ..., f_N$ are independent, this procedure leads to a vector of restricted maximum likelihood estimates for probabilities $p_1, p_2, ..., p_N$. Such estimates may find non-trivial applications in various fields of study including medicine, toxicology or calculating insurance premiums ([14],[16]). At the same time, for most sampling schemes a significant dependence between some sample membership indicators may appear. Resulting lack of independence among $f_1, ..., f_N$ seems to prevent the use of PAVA to estimate ordered inclusion probabilities. However, new results obtained by Gamrot [7] indicate that PAVA-based estimates remain consistent even in the case of strong correlation between sample membership indicators. Hence the empirical Horvitz-Thompson estimator constructed upon them should also remain consistent. A more detailed investigation of its properties for a specific sampling scheme is presented in the next section.

3 Simulation results

The sequential fixed-cost sampling scheme of Pathak [11] features varying inclusion probabilities. Their exact evaluation is very demanding computationally even for modest sample sizes. Nevertheless, despite the existence of sufficiency-based unbiased estimators that do not rely on inclusion probabilities, empirical H-T estimation may be of interest when nonresponse corrections need to be incorporated or when some modifications are introduced to the original scheme. In this section the Pathak procedure in its original form serves as an illustration of PAVA-based empirical H-T estimation. The selection procedure works in the following way. Let $c_1, ..., c_N$ be costs of observing the value of characteristic under study for corresponding population units, known in advance. Without a loss of generality one may assume that units are pre-ordered by decreasing value of this cost so that $c_1 \geq c_2 \geq ... \geq c_N$. Individual units are drawn to the sample one-by-one with equal probabilities until the sum of costs corresponding to drawn units exceeds some pre-determined survey budget C. More specifically, the procedure is stopped when the cumulative cost of the sample becomes greater or equal C and the population unit for which it occurs is not included in the sample. As a result, inclusion probabilities satisfy the simple order expressed by multiple inequality (4). Hence their estimates $\hat{p}_1, ..., \hat{p}_N$ may be obtained through PAVA. By inserting these estimates into formula (2) an estimator of the population total for the Pathak sampling scheme is obtained.

A simulation study was carried out in order to assess the properties of resulting H-T estimator for the population total and to compare it to the classic empirical H-T statistic involving inclusion probability estimates computed through Fattorini's formula (3). In simulation experiments, a sampling frame corresponding to the finite population under study was represented by the data set obtained during agricultural census carried out by Polish Central Statistical Office (GUS) in 1996. The dataset described population of 695 farms in the Gręboszów municipality of the Dąbrowa Tarnowska district. Total yearly sales of a farm represented the variable under study for which the population total was to be estimated. It was also assumed that the cost of observing this variable for individual farms was proportional to the farm area, assumed to be known. It was assumed that $C = 0.05 \cdot (c_1 + ... + c_N)$ so the survey budget was equal to five percent of the census cost. As a result sample size could vary in the range between 10 and 102 units depending on the cost of sampled units. The simulation study was designed to jointly capture the variability of estimates resulting from both sampling of finite population and a random simulation study. It was carried out by sequentially generating independent sets of replications for each sample drawn from the finite population and computing corresponding empirical H-T estimates. All computations were made in the R computing environment [12].

In the first experiment $R = 300$ replications were drawn for each of 15000 Pathak samples. Histograms of empirical distributions for both empirical H-T estimates are shown in the Figure 1. The dashed vertical line represents true value of the population total while the solid vertical line represents observed average of estimates. Moreover, observed characteristics of empirical distributions for both empirical H-T estimators are listed in the Table 1. They include their biases, relative biases (Relbias) computed as a ratio of bias to the true estimated value, variances, mean square errors (MSE), relative root mean square errors (RRMSE) and ratios of squared bias to the total mean square error (Bias share). Distributions of both estimators feature a very similar shape, with slight positive skew and nearly the same variance. However the distribution of Fattorini's estimator is substantially shifted to the left which is reflected by its strong

negative bias. The bias of the PAVA-based estimator is positive, but its absolute value is several times lower. This advantage in terms of absolute bias also influences the overall accuracy of estimates. The mean square error of the PAVA-based estimator is 17 % lower than that of Fattorini's statistic.

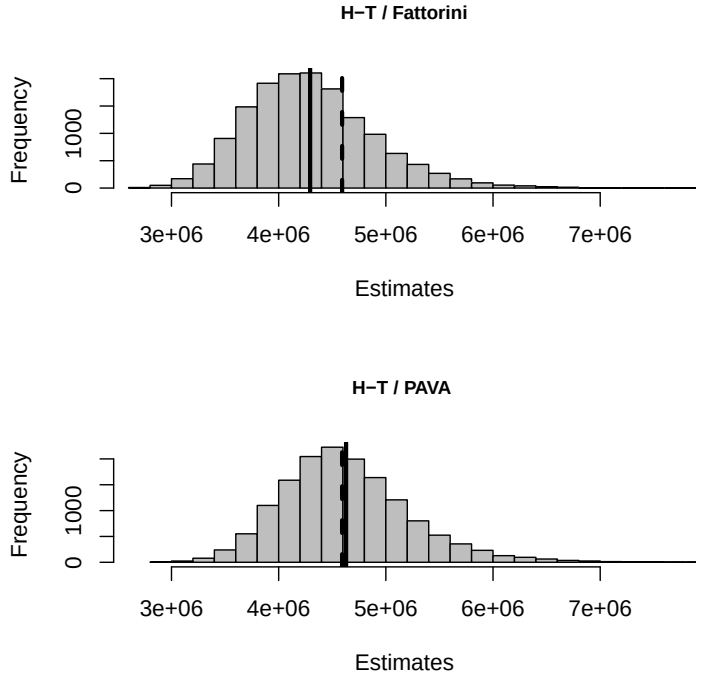


Figure 1: Distribution of estimates for Fattorini's and PAVA-based empirical H-T statistic

	Bias	**Relbias**	**Variance**	**MSE**	**RRMSE**	**Bias share**
Fattorini	-298242.8	-0.0649	$3.592 \cdot 10^{11}$	$4.482 \cdot 10^{11}$	0.1457	0.1984
PAVA-based	37207.4	0.0081	$3.702 \cdot 10^{11}$	$3.716 \cdot 10^{11}$	0.1327	0.0037

Table 1: Selected distribution characteristics of both empirical H-T estimators

In the second experiment, the investigation was extended to compare the behavior of estimators for $R = 100, 200, ..., 1000$ replications. For each value of R a total of 10000 samples were drawn using Pathak scheme, with corresponding set of R replications again generated independently for each sample. The absolute bias, relative bias, relative root mean square error and the share of bias in the MSE for varying values of R are shown in the Figure 2. It turns out that the bias of the proposed PAVA-based estimator is very stable for small numbers of replications, while it slowly tends to zero when R grows. For Fattorini's estimator this tendency was more pronounced, but absolute values of bias were 13 to 3.46 times higher reaching nearly 20 percent for $R = 100$. The relative root mean square error of both estimators exhibits similar behavior. For Fattorini's statistic it is always higher than for the PAVA-based one, with the relative difference reaching 58 percent for $R = 100$ but quickly diminishing when R grows. For any value of R the share of bias in the mean square error did not exceed one percent for the PAVA-based estimator while it reached over 70 percent for the Fattorini's statistic and $R = 100$.

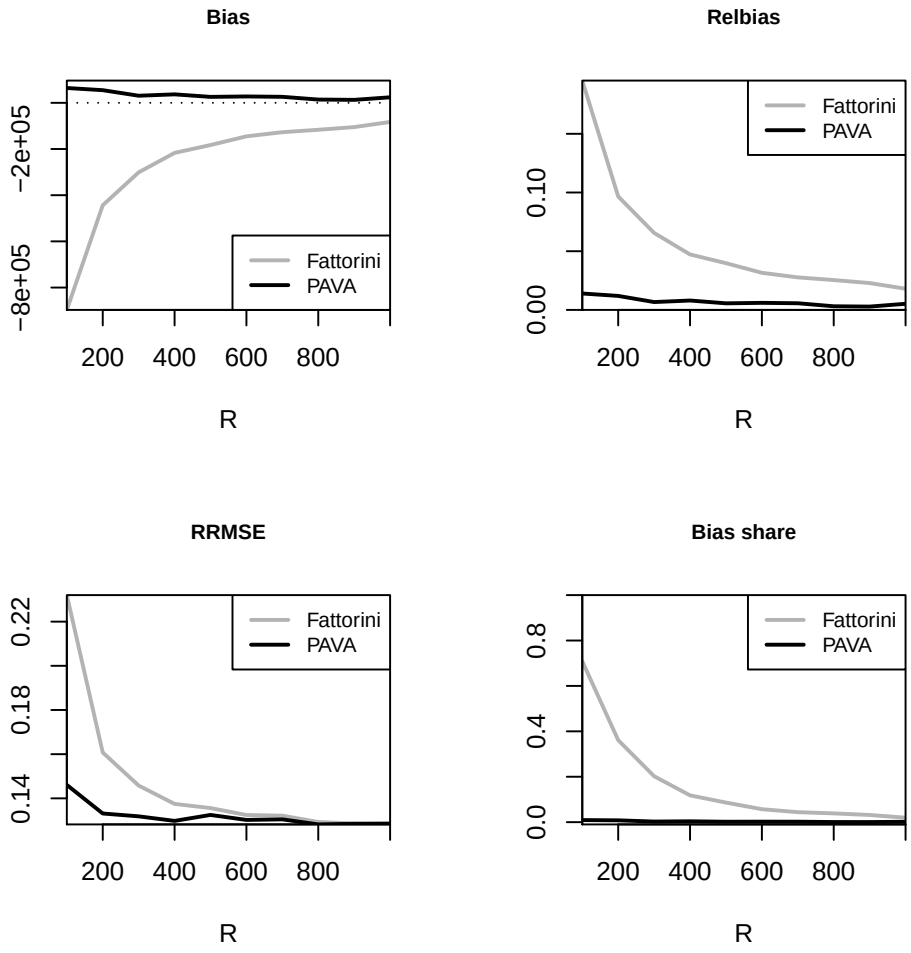


Figure 2: Stochastic properties of Fattorini's and PAVA-based empirical H-T statistic for varying R

4 Conclusions

The proposed estimator indirectly utilizes observations of sample membership indicators associated with all population units to increase the accuracy of first-order inclusion probability estimates corresponding to sampled units. Presented simulation study suggests, that this in turn significantly reduces the bias and improves the accuracy of empirical H-T estimator itself, even for quite large numbers of replications. One may reasonably expect that the strength borrowing effect should be particularly beneficial when differences between true individual inclusion probabilities are small, the population size is large and when generation of sample replications is time-consuming.

Acknowledgements

The work was supported by the grant No N N111 558540 from the Ministry of Science and Higher Education

References

[1] Ayer, M., and Brunk, H.D., and Ewing, G.M., and Reid, W.T., and Silverman, E.,: An empirical Distribution function for Sampling with Incomplete Information, *The Annals of Mathematical Statistics* **6** (1955), 641-647.

[2] Barabesi, L., and Fattorini, L., and Ridolfi, G.: Two-phase surveys of elusive populations. In *Proceedings of the Statistics Canada Symposium 97: New Direction in Surveys and Censuses*, Statistics Canada, Ottawa 1997, 285-287.

[3] de Leeuw, J., and Hornik, K., and Mair, P.: Isotone optimization in R: Pool-adjacent-violators algorithm (PAVA) and active set methods, *Journal of Statistical Software* **32** (2009), 1-24.

[4] Fattorini, L., and Ridolfi, G.: A sampling design for areal units based on spatial variability, *Metron* **55** (1997): 59-72.

[5] Fattorini, L.: Applying the Horvitz-Thompson criterion in complex designs: A computer-intensive perspective for estimating inclusion probabilities, *Biometrika* **93** (2006), 269-278.

[6] Fattorini, L.: An adaptive algorithm for estimating inclusion probabilities and performing the Horvitz-Thompson criterion in complex designs, *Computational Statistics* **24** (2009), 623-639.

[7] Gamrot, W.: On Pool-Adjacent-Violators Algorithm and its Performance for Non-Independent Variables, To appear in *Studia Ekonomiczne* (2012).

[8] Härdle, W.,: *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, 1992.

[9] Horvitz, D.G., and Thompson, D.J.,: A Generalization of Sampling Without Replacement from a Finite Universe, *Journal of the American Statistical Association* **47** (1952), 663-685.

[10] Myrskylä, M.: Generalized Regression Estimation for Domain Class Frequencies. Research report, Statistics Finland, Helsinki 2007.

[11] Pathak, P.K.,: Unbiased estimation in fixed-cost sequential sampling schemes, *Annals of Statistics* **4** (1976), 1012-1017.

[12] R Development Core Team: A language and environment for statistical computing (2011), R Foundation for Statistical Computing, Vienna.

[13] Rosén, B.: On sampling with probability proportional to size, *Journal of Statistical Planning and Inference*, **62** (1997), 159-191.

[14] Stylianou, M., and Proschan, M., and Flournoy, N.,: Estimating the probability of toxicity at the target dose following an up-and-down design, *Statistics in Medicine* **22** (2003), 535-543.

[15] Thompson, M.E., and Wu, C.: Simulation-based randomized systematic PPS sampling under substitution of units, *Survey Methodology* **34** (2008), 3-10.

[16] Wolny-Dominiak, A.: The multi-level factors in insurance rating technique. In: *Proceedings of 27th international conference "Mathematical Methods in Economics 2009"* (Brožová, H. ed.), Czech University of Life Sciences, Prague 2009, 346-351.

Matrix period in max-drast fuzzy algebra

Martin Gavalec[1], Zuzana Němcová[2]

Abstract. Periods of matrix power sequences in max-drast fuzzy algebra and methods of their computation are considered. Matrix power sequences occur in the theory of complex fuzzy systems with transition matrix in max-t algebra, where t is a given triangular fuzzy norm. Interpretation of a complex system in max-drast algebra reflects the situation when extreme demands are put on the reliability of the system. For triangular norm $t = \min$, the matrix periods have been already described in literature. In this paper, a polynomial algorithm for computing the matrix period in max-drast algebra is presented, and the relation between matrix periods in max-min and max-drast cases is described.

Keywords: drastic triangular norm, max-drast algebra, matrix powers, fuzzy systems.
AMS classification: 08A72, 90B35, 90C47

1 Introduction

Transition matrices with fuzzy values are applied in fuzzy approach to discrete time systems. Various triangular fuzzy norms t are used, in dependence on the character of the system. Development of the system in time is then described by power sequence of the transition matrix, computed in max-t fuzzy algebra. Steady states of the system correspond to periodic behavior of the power sequence.

Max-t fuzzy algebra uses instead of conventional operations for multiplication of vectors and matrices the operations of max and one of the triangular norms - min, drast, prod or Lukasziewicz. Algebras using operations maximum and minimum, possibly in combination with addition or multiplication, were studied by many authors, see e.g. [2], [3], [4], [5], [6], [8], [9], [10]. For triangular norm $t = \min$, the matrix periods are described in [7].

Matrix power sequences and their periodicity in max-drast algebra are studied in this paper and a polynomial method for computation of the matrix period is described. New elements do not appear in the matrices during the computation of matrix powers by operation max and drast. Therefore, some repetition in the matrix power sequence always occurs, which inevitably leads to periodic behavior.

The drastic triangular norm is the basic example of a non-divisible t-norm on any partially ordered set, see [1]. Interpretation of a complex discrete time system in max-drast algebra reflects the situation when extreme demands are put on the reliability of the system. The matrix powers in max-drast algebra behave differently than those in max-min algebra, however the properties of matrix periods in both cases show some similarity described in the paper.

2 Basic notions

In the paper we work with the max-drast fuzzy algebra $(\mathcal{I}, \oplus, \otimes_d)$, where $\mathcal{I}$ is the unit interval $\langle 0, 1\rangle$, $\oplus$ is the maximum operation on $\mathcal{I}$ and $\otimes_d$ is the binary drast operation (drastic triangular norm) on $\mathcal{I}$ defined as follows

$$\mathrm{drast}(x,y) = \begin{cases} \min(x,y) & \text{if } \max(x,y) = 1 \\ 0 & \text{if } \max(x,y) < 1 \end{cases}$$

[1]University of Hradec Králové/Faculty of Informatics and Management, Rokitanského 62, Hradec Králové 3, 500 03, e-mail: martin.gavalec@uhk.cz
[2]University of Hradec Králové/Faculty of Informatics and Management, Rokitanského 62, Hradec Králové 3, 500 03, e-mail: zuzana.nemcova@uhk.cz

Let n be a fixed natural number, we shall use the notation $N = \{1, 2, \ldots, n\}$. Let $A \in \mathcal{I}(n, n)$ be a fixed matrix, the entries of A are denoted as $a_{ij} \in \mathcal{I}$, $i, j \in N$.

If A is interpreted as the reliability matrix of transitions in a discrete time fuzzy system, then the drast norm in the computation of the matrix powers reflects extreme demands on reliability in the system. This follows from the computation of the powers of A. More precisely, zero element $a_{ij} = 0$ of the matrix A means that the reliability of the transition from state i to state j in the system is zero. If $a_{ij} = 1$, then the transition between states i and j is reliable. On the other hand, if value of a_{ij} lies between zero and one, i.e. $a_{ij} = u$, $u \in (0, 1)$, then some uncertainty in the transition expressed by the value u must be considered. The smaller value, the higher degree of uncertainty is expressed.

The idea of reliability of edges (one-step transitions) can be extended to the reliability of paths in the characteristic weighted digraph $G(A)$ of matrix A (several-step transitions). The nodes of $G(A)$ correspond to states of the system and are denoted by numbers in N, while an edge from node i to node j has the weight $w(i, j) = a_{ij}$, for all $i, j \in N$. Getting from state i to state j reliably means that all necessary steps are reliable (weighted by 1). If there is one uncertain step (having positive weight smaller than 1) in the sequence of steps, then the sequence is considered as uncertain. Extreme demands on the reliability of the system imply that any sequence containing two or more uncertain steps is considered as inadmissible.

Formally, a path is called *strong* if all edges of the path are reliable (weighted by 1). The path is called *weak* if and only if exactly one edge of the path is weighted by $u \in (0, 1)$ and the weight of the remaining edges is 1. Finally, the path is called *inadmissible* if more than one edge in the path is weighted by value less than 1, in other words, there is more than one uncertainty on the path from node i to j, or if all edges in the path have zero weight. The set of all strong (weak) paths is denoted by SR (WR). The notation $SR(i, j, r)$ ($WR(i, j, r)$) stands for the set of all paths in SR (WR) from i to j of length r.

3 Matrix powers in max-drast algebra

Definition 1. Let $i, j \in N$, let $p = (i_0, i_1, \ldots, i_r)$, with $i_0 = i, i_r = j$ be a path in the $G(A)$. The weight of path p in max-drast algebra is set to the drastic fuzzy product

$$w_d(p) := \bigotimes_d \{a_{i_{s-1}i_s};\ s = 1, 2, \ldots r\} \ . \tag{1}$$

Let us denote the entries of the r-th power, A^r, of matrix A as a^r_{ij} . The interpretation of this value is given in the following proposition.

Proposition 1. *Let $A \in \mathcal{I}(n, n)$, $i, j \in N$, let r be a natural number. Then the entries of A^r have values given by the formula*

$$a^r_{ij} = \bigoplus \{w_d(p);\ p \text{ is a path in } G(A) \text{ from node } i \text{ to node } j \text{ of length } r\} \ . \tag{2}$$

Proof. The assertion is proved easily by recursion on r, using Definition 1 and the matrix multiplication in max-drast algebra. □

More specific formula for the values of elements in r-th matrix power in max-drast algebra is presented in the next proposition.

Proposition 2. *Let $A \in \mathcal{I}(n, n)$, $i, j \in N$, let r be a natural number. Then the entries of A^r are described by the formula*

$$a^r_{i,j} = \bigoplus \{w_d(p);\ p \in SR(i, j, r) \cup WR(i, j, r)\} \ . \tag{3}$$

Proof. The assertion follows from Definition 1 and Proposition 1. Clearly, by the definition of the operation $\otimes_d$, the weight of every inadmissible path p is zero. If the set in equation (3) is empty, then the result of the operation $\bigoplus$ is equal to zero, which corresponds to the fact that there are neither strong nor weak paths from i to j in this case, while the inadmissible paths have zero weights. □

Proposition 3.

$$w_d(p) > 0 \quad \text{if and only if} \quad p \in SR(i,j,r) \cup WR(i,j,r) \tag{4}$$

$$a^r_{i,j} > 0 \quad \text{if and only if} \quad \exists p \in SR(i,j,r) \cup WR(i,j,r) \tag{5}$$

Proof. The assertion follows directly from previous considerations. □

The last proposition in this section describes the conditions under which the weight of a path in max-drast is equal to 1 or to some value u with $0 < u < 1$.

Proposition 4. *For weight of a path p between nodes i and j of length r the following assertions hold true*

$$w_d(p) = 1 \quad \text{if and only if} \quad p \in SR(i,j,r) \tag{6}$$

$$0 < w_d(p) < 1 \quad \text{if and only if} \quad p \in WR(i,j,r) \tag{7}$$

Proof. This proposition results from the definition of strong and weak path. Weights of all edges in strong path are 1, hence the weight of such path is 1. Similarly, the weight of the weak path (with exactly one weak edge and remaining strong edges) is the least element $u \in (0,1)$. □

In the next section we shall consider subdigraphs of $G(A)$, called *threshold digraphs*, which will be denoted as $G(A,h)$, where $h \in \langle 0,1\rangle$ is a threshold level. In the threshold digraph $G(A,h)$, such edges are only included, whose weight is equal or greater than level h. We also shall consider strongly connected components of $G(A,h)$. By standard definition, strongly connected component K is such a subset of nodes in the digraph that any two nodes i, j are contained in a common cycle.

Component K is called non-trivial if there is at least one cycle of positive length in K. The set of all non-trivial strongly connected components of digraph $G(A,h)$ will be denoted by $\mathrm{SCC}^*(G(A,h))$. For $K \in \mathrm{SCC}^*(G(A,h))$, the component period $\mathrm{per}(K)$ is defined as the greatest common divisor of the length of all cycles in K.

4 Computation of the matrix period

The max-min and max-drast algebra differ in the interpretation. A max-min matrix describes the flow capacities, while a max-drast matrix concerns the transitions reliability in the system.

Similarly as in the max-min algebra, by max-drast operations no new elements (except 0) are created. As a consequence, the matrices in the power sequence of matrix A in max-drast algebra only contain the entries from A.

The above mentioned reliability of the path is represented by powers of the matrix A. We can investigate paths of length that is even much longer than number of edges in the digraph of A. This means to compute high powers of matrix A. As a consequence of repetition of elements and possibility of creation only the zero element in the matrix, the period in the power sequence of A, $\mathrm{per}(A)$, must sooner or later occur.

Computation of the length of the matrix period in max-min algebra has been described in [7]. The matrix period in max-min algebra is computed by the following formula.

$$\mathrm{per}_{\min}(A) = \mathrm{lcm}\big\{\mathrm{per}(K);\, K \in \mathrm{SCC}^*(G(A,h)), h \in \langle 0,1\rangle\big\} \ . \tag{8}$$

In computation of the matrix period in max-drast algebra, only the level $h = 1$ must be considered.

Theorem 5. *Let $A \in \mathcal{I}(n,n)$. Then*

$$\mathrm{per}_{\mathrm{drast}}(A) = \mathrm{lcm}\big\{\mathrm{per}(K);\, K \in \mathrm{SCC}^*(G(A,1))\big\} \ . \tag{9}$$

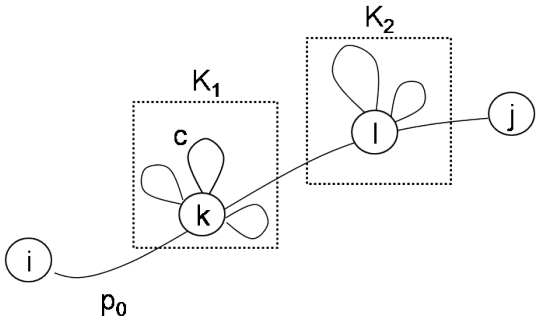


Figure 1: The path p_0

Proof. Let us denote $d^* = \operatorname{lcm} \operatorname{per}(K)$, where $K \in \mathrm{SCC}^*(G(A,1))$. We also denote the period of the matrix A in max-drast algebra as $d_A = \operatorname{per}_{\mathrm{drast}}(A)$. It is the least d such that $\exists R, \forall r > R : A^r = A^{r+d}$. Determine the period of the elements of the matrix powers as $d_{i,j} = \operatorname{per}_{\mathrm{drast}}(A_{i,j})$. This is also the least d such that $\exists R, \forall r > R : a^r_{i,j} = a^{r+d}_{i,j}$. Let us denote $d_A = \operatorname{lcm}\{d_{ij}, i,j \in N\}$. In this notation the assertion of the theorem can be simply written as $d_A = d^*$.

The proof will be done in two steps. In step (i) we prove that $d_A | d^*$, and in step (ii) that $d^* | d_A$.

(i) It holds that $\operatorname{lcm} d_{i,j} | d^*$ if and only if $(\forall i,j) : d_{i,j} | d^*$ - it is clear, that if least common multiple (e.g. $\operatorname{lcm} = a \cdot b$) divides some number, then also particular component of the multiple (e.g. a) divides the number. Consider non-trivial $K \in \mathrm{SCC}^*(G(A,1))$ and $p_0 \in SR(i,j,r) \cup WR(i,j,r)$ is a path of length r between i and j which goes through the strongly connected components K (see Figure 1).

Then we can find such $a^r_{i,j}(p_0)$ that equals to some $a^{r+d}_{i,j}(p_0)$. It holds that weight of such path is $w(p_0 + C) = w(p_0) \otimes_d w(C) = w(p_0)$, where C is certain combination of strong cycles in the component $C \subseteq K$. Because $d_{i,j}(p_0)$ is a period dependent on great common divisor of K, we can write that $d_{i,j}(p_0) | \operatorname{per}(K) | d^*$. Element of the r-th power of A is then $a^r_{i,j} = \max\{a^r_{i,j}(p_0)\}$. If $d_{i,j}(p_0) | d^*$ then also $d_{i,j} | d^*$. Let us denote $d^*_{i,j} = \operatorname{lcm} d_{i,j}(p_0)$. Let d be such that $\forall p_0 : a^r_{i,j}(p_0) = a^{r+d}_{i,j}(p_0)$ then also $a^r_{i,j}(p_0) = a^{r+d^*_{i,j}}_{i,j}(p_0)$ and we can write also that $a^r_{i,j} = a^{r+d^*_{i,j}}_{i,j}$ As we proved, $d_{i,j} | d^*$. $d^*_{i,j} = \operatorname{lcm} d_{i,j}(p_0)$ and thus $d^*_{i,j} | d^*$. This results in $d_{i,j} | d^*_{i,j} | d^*$.

(ii) We have to prove that $d^* | d_A$, i.e. $\operatorname{lcm}\big\{\operatorname{per}(K), K \in \mathrm{SCC}^*(G(A,1))\big\} | d_A$. Let $K \in \mathrm{SCC}^*(G(A,1))$, $i \in [K]$. We claim, that $\operatorname{per}(K) | d_{i,i}$ which implies $\operatorname{per}(K) | d_A$, i.e. $d^* | d_A$.

Proof of the claim: Let d be such that $a^r_{i,i} = a^{r+d}_{i,i}$ for all $r > R$. Then take such r, that $\operatorname{per}(K) | r$, then also $\operatorname{per}(K) | r + d$ and its clear that also $\operatorname{per}(K) | d$. As $d_{i,i}$ is such d, that $a^r_{i,i} = a^{r+d}_{i,i}$, then we have $\operatorname{per}(K) | d_{i,i}$. And $d_A = \operatorname{lcm} d_{i,i}$, hence $\operatorname{per}(K) | d_{i,i} | d_A$. □

From the computation of the drast operation we can see, that zero is the result of the case where $\max(x,y) < 1$. This result is thus insensitive to the variability of the elements x and y, which are less than one.

Theorem 6. *Let $A \in \mathcal{I}(n,n)$. If all elements of the matrix A, denoted as $u \in (0,1)$, are replaced by one constant value $c \in (0,1)$, then the period of matrix power in max-drast will not be changed.*

Proof. Reliability of the path of r steps from node i to node j, $a^r_{i,j}$, is computed as maximum of paths weighted by drast operation. It holds, that the element $a^r_{i,j}$ is equal to value of the most reliable path of length r. For computation of d, the period of the element $a_{i,j}$, consider the $p_0 \in WR(i,j,r) \cup SR(i,j,r)$ which goes through the strongly connected components K (see Figure 1). If $p_0 \in WR(i,j,r)$, then weight of the path is changed from u to c without the change in the period of the element. On the other side, if the path $p_0 \in SR(i,j,r)$ then two possibilities can arise. The first is that for computation of the element $a^{r+1}_{i,j}$ is used strong cycle, then $a^r_{i,j} = a^{r+1}_{i,j} = 1$. The second is that for computation of the element $a^{r+1}_{i,j}$ is used weak cycle and $a^{r+1}_{i,j} = c$. In any case at least of one element $a^{r+l}_{i,j}, l < d$ will be equal to 1, therefore the period can not be changed. □

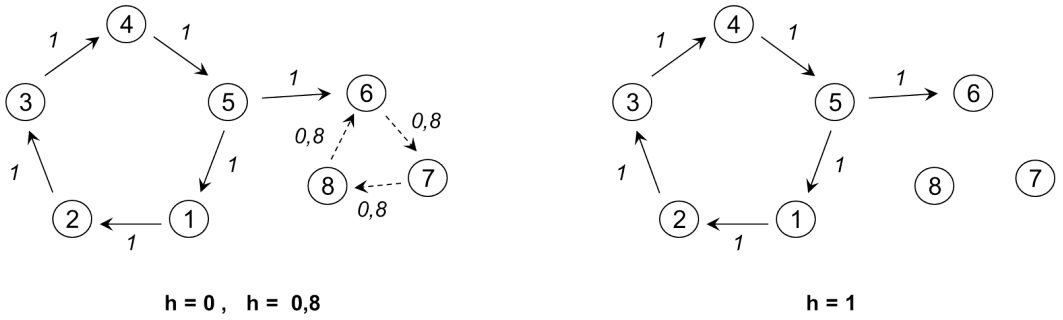


Figure 2: Treshold graphs

Theorem 7. *Let $A \in \mathcal{I}(n,n)$. Matrix power periods in max-min and max-drast equals if and only if edges on level $h \in u$ do not create new component C, such that*

$$\operatorname{lcm}\{\operatorname{per}(K);\, K \in \operatorname{SCC}^*(G(A,1))\} \,|\, \operatorname{per}(C) \ . \tag{10}$$

Proof. This results from the computation of the matrix powers periods in both algebras. In max-drast algebra only the level $h = 1$ must be considered. While in max-min algebra, periods of all non-trivial strongly connected components, K all levels can influence matrix power period by final computation of the least common multiple of $\operatorname{per}(K)$ through all levels. It is clear that by increase of the quantity of cycles in some component K_i, the greatest common divisor of this cycles can decrease. The values of periods differ in only one case, when on some level $h \neq 1$ appears such K_i, whose $\operatorname{per}(K_i)$ is not the divisor of the matrix period in max-drast, therefore this $\operatorname{per}(K_i)$ must be included into the computation of the least common multiple and will change the matrix power period in max-min. □

Theorem 8. *Let $A \in \mathcal{I}(n,n)$. Then it holds that*

$$\operatorname{per}_{\text{drast}}(A) | \operatorname{per}_{\min}(A) \ . \tag{11}$$

Proof. The calculation of matrix period in both algebras differs in final computation of least common multiple. Hence it holds, that the period in max-drast is a divisor of period in max-min. □

5 Example

Theorem 7 shows the case when period in max-min and max-drast equals. The following example (see the matrix A below and corresponding threshold graphs - Figure 2) represents the case when the period computed in both algebras differs.

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,8 \\ 0 & 0 & 0 & 0 & 0 & 0,8 & 0 & 0 \end{pmatrix} \tag{12}$$

Computation of the period in max-min algebra covers all threshold levels $h = \{0,8; 1\}$. At level $h = 0,8$ we can identify two non-trivial strongly connected components $K_1 = \{1,2,3,4,5\}$ and $K_2 = \{6,7,8\}$. Period of the first component, K_1, is equal to greatest common divisor of cycles in this component. Because the cycle is the only one, $\operatorname{per}(K_1) = 5$. Similarly, $\operatorname{per}(K_2) = 3$. At level $h = 1$ there is only one strongly connected component $K_3 = \{1,2,3,4,5\}$, $\operatorname{per}(K_3) = 5$. Then matrix power period in max-min algebra is computed as $\operatorname{per}_{\min}(A) = \operatorname{lcm}\{\operatorname{per}(K_1), \operatorname{per}(K_2), \operatorname{per}(K_3)\} = \operatorname{lcm}\{5,3,5\} = 15$.

In max-drast algebra only the $h = 1$ must be considered - there is exactly one non-trivial strongly connected component at this level, $K_4 = \{1, 2, 3, 4, 5\}$. Period of this component is also the greatest common divisor of cycles in the component: $\mathrm{per}(K_4) = 5$, then $\mathrm{per}_{\mathrm{drast}}(A) = \mathrm{lcm}\{\mathrm{per}(K_4)\} = \mathrm{lcm}\{5\} = 5$.

Acknowledgements

The support of INKOV project # CZ.1.07/2.3.00/20.0001 financed from EU and Czech Republic funds, and Czech Science Foundation project # 402/09/0405 is kindly acknowledged.

References

[1] Casasnovas, J., Mayor, G.: Discrete t-norms and operations on extended multisets. *Fuzzy Sets and Systems* **159** (2008), 1165–1177.

[2] Baccelli, F. L., and Cohen, G., and Olsder, G. J., and Quadrat, J. P.: *Synchronization and Linearity. An Algebra for Discrete Event Systems.* Wiley, Chichester, 1992.

[3] G. Cohen, D. Dubois, J. P. Quadrat, M. Viot, *A linerar-system-theoretic view of discrete event processes and its use for performance evaluation in manufacturing*, IEE Transactions on Automatic Control, AC-30 (1985), 210-220.

[4] R. A. Cuninghame-Green, *Describing industrial processes with interference and approximating their steady-state behavior*, Oper. Res. Quart. **13** (1962), 95-100.

[5] R.A. Cuninghame-Green: *Minimax Algebra*, Lecture Notes in Economics and Mathematical Systems **166**, Springer–Verlag, Berlin, 1979.

[6] R.A. Cuninghame-Green: *Minimax Algebra and Application*, in: Advances in Imaging and Electron Physics **90**, (P. W. Hawkes, Ed.). Academic Press, New York, 1995.

[7] Gavalec, M.: Computing matrix period in max-min algebra, *Discrete Appl. Math.*, **75** (1997), 63–72.

[8] Robertazzy, T.G.: *Planning Telecomunication Networks.* IEEE: Piscataway, NJ. 1999.

[9] Yi-Jia Tan, *On the powers of matrices over a distributive lattice*, Lin. Algebra Appl. **336** (2001), 1-14.

[10] U. Zimmermann: *Linear and Combinatorial Optimization in Ordered Algebraic Structure*, Ann. Discrete Math. **10**, North Holland, Amsterdam, 1981.

Relevance of the material deprivation indicator, evidence based on Slovak EU-SILC microdata

Roman Gavuliak[1]

Abstract. The indicator of material deprivation is defined as people living in households fulfilling four out of nine conditions defined in the EU SILC methodology. It is also a vital area of interest for the recently created Europe 2020 strategy. In this paper we analyse the construction of the material deprivation strategy validating its informational value as well as test for its validity in case of adding additional dimensions of material deprivation through the use of logistic regression. As a conclusion we propose a change to construction of the indicator.

Keywords: Material deprivation, EU SILC, Europe 2020, Indicator construction

JEL Classification: I32
AMS Classification: 62P25

1 Material deprivation

Material deprivation is considered as one of the dimensions of social exclusion. The discourse about social exclusion originates from France [6] and can the term can be defined in a number of ways. Room [8] characterized social exclusion as a situation where (a) individuals suffer generalized disadvantage in terms of education, training, employment, housing, financial resources, etc; (b) individuals' chances of gaining access to the major social institutions which distribute these life chances are substantially less than those of the rest of the population; (c) these disadvantages persist over time.

Material deprivation, while dependent on the income of an individual is more focused on the non-financial possibilities and resources avaiable to the individual. The notion for the creation of a **common EU-wide measure** of material deprivation first appeared during the design phase of the **Lisbon Strategy** [1]. Several member states had however already constructed and measured material deprivation. Table 1 below illustrates the measures implemented by one or more member states of the EU before the year 2000.

Indicator	Countries using the indicator
Percentage of households lacking specified amenities	Belgium, Spain, France, Ireland, Italy, Portugal, Finland, United Kingdom
Number of homeless	Belgium, Spain, France, Ireland, United Kingdom
Amount of people who have had difficulties in the last 12 months in paying for water, electricity, and gas	France
Percentage of population in rent arrears	France
Percentage of population not going away on holiday	France
Measure of 'consistent' poverty (a combination of financial and non-financial poverty)	Ireland
Percentage with a subjective perception of difficulty of access to some selected services (medical, food stores, schools, police stations, etc.)	Italy
Percentage of population living in housing that does not meet decency standards	United Kingdom
Percentage of households with children living in temporary accommodation	United Kingdom

Table 1 Material deprivation indicators in EU member states, before 2000

[1] University of Economics in Bratislava, Department of Statistics, Dolnozemská cesta 1/b 852 35 Bratislava Slovakia, roman.gavuliak@gmail.com

1.1 Material deprivation within the EU context

The EU concept of measurement [1] of material deprivation is supposed to supplement the imperfections accompanying measures of financial poverty. **Indicators of deprivation are generally based on asking survey respondents whether they have a specific item or participate in a certain activity**. The commission responsible for the development of the indicator in regards to material deprivation recommended developing the following measurements: non-monetary indicators of deprivation, Housing of poor environmental quality, housing costs, homelessness and precarious housing

The first EU-wide indicator of material deprivation was developed as a part of the EU Statistics about income and living conditions (EU SILC), which was first conducted in 2003. The first round of EU SILC conducted in Slovakia occurred in 2005 one year after the entry into the EU. The material deprivation indicator however wasn't introduced as a part of the Lisbon Strategy (which was terminated due to its failure in 2009). It appeared in 2009 as a part of the **Sustainable Development Strategy** of the EU [4] within the context of social inclusion [2]. The official definition [5] of the indicator describes it as covering issues relating to economic strain and durables. **Severely materially deprived persons have living conditions greatly constrained by a lack of resources and cannot afford at least four of the following**: to pay rent or utility bills; to keep their home adequately warm; to pay unexpected expenses; to eat meat, fish or a protein equivalent every second day; a week holiday away from home; a car; a washing machine; a colour TV; or a telephone. Table 2 illustrates the development of the material deprivation levels for Slovakia.

Year	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
Value	22.1	18.2	13.7	11.8	11.1	11.4

Table 2 Levels of material deprivation in Slovakia, 2005-2010

As we can observe, levels of material deprivation in Slovakia have been experiencing a decelerating decrease until 2009, followed by a slight increase due to the impact of the global economic and budget deficit crisis.

European Commissions has, as the reaction to the failure of the Lisbon strategy and in the light of the ongoing global economic crisis created strategy **Europe 2020**. Europe 2020 is supposed to help EU cope with the current economic environment and increase its competitive potential by 2020. Europe 2020 also includes goals for the area of social inclusion. **One of the main indicators used to monitor the progress in this area is the not by a coincidence the material deprivation indicator we described above**.

2 Assessment of the indicator

For evaluating the relevance of the material deprivation indicator we will be using the EU SILC 2010 data spanning 5376 households comprising of 16304 persons. It should be noted, that while the indicator expresses the percentage of materially deprived individuals on the whole population of an EU member state, the access to different amenities is measured at a household level. There are two ways of conducting the analysis of the indicator.

We can **analyse** changes in its values in regards to the threshold formed by the **number of conditions met in order to classify an individual as materially deprived**, this involves basic computation methods of summation and evaluation of logical conditions.

The second option is analysing the conditions themselves, either by **assessing the significance of the conditions** already implemented as a part of the indicator or by taking a look at conditions that could possibly be implemented (based on data availability). In order to evaluate the statistical significance of the conditions in regards to whether an individual is considered to be materially deprived, we will be **using logistic regression**.

In this paper we will use a combination of both methods in order to take into consideration both the informational value of the indicator as a whole as well as the significance of the conditions it includes.

Logistic regression

Logistic regression [7] statistical method similar to multiple regression. It addresses the issue of having a dichotomous dependent variable. In our case the dependent variable is, whether an individual within the sample is or isn't materially deprived. The difference compared to multiple regression is way of estimating its variates – by applying a logit transformation to the dependent variable. This creates several differences within the estimation process as well as the interpretation of the explanatory variables.

The coefficients for independent variables are estimated using maximum likelihood method and will results in two kinds of values – logit values or an odds value.

The original logistic coefficients (logit) measure the direction of the relationship between the individual explanatory variables and the dependent variable. We estimate them using (1):

$$Logit_i = \ln(\frac{prob_{event}}{1 - prob_{event}}) = b_0 + b_1 x_1 + ..b_0 + b_n x_n \quad (1)$$

A second possibility is estimating the exponentiated logistic coefficients (odds value) as shown in (2). Odds values describe the magnitude of the relationship between the explanatory and the dependent variable. It should be noted [7] that both formulations are equivalent. In our analysis we estimate both the logistic coefficients and their exponentiated form.

$$Odds_i = (\frac{prob_{event}}{1 - prob_{event}}) = e^{b_0 + b_1 x_1 + ...b_0 + b_n x_n} \quad (2)$$

As for the model fit, we will be using two main Pseudo R^2 measures - Cox and Snell R2 and Nagelkerke R^2, both operating similarly to basic R^2 measures, with Naegelkerke R^2 being a more accurate measure [9].

2.1 Changes in the amount of conditional requirements

In this part we will see how the values of the indicator change based on the number of conditions that have to be met; we illustrate these changes in Figure 1. Of course we can expect that decreasing the number of conditions that need to be met in order for a person to be materially deprived will feature the increase of persons classified as materially deprived, while an increase of conditional requirements is prone to trigger a decrease.

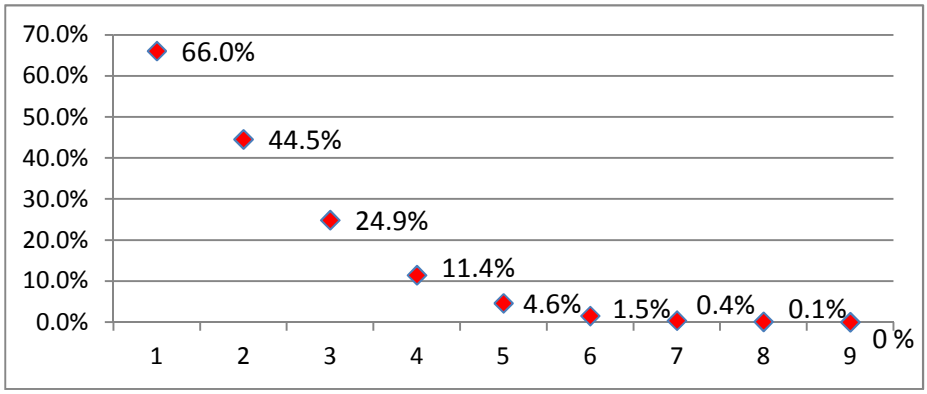


Figure 1 Levels of material deprivation based on the number of conditions which have to be met, Slovakia 2010

Judging from Figure 1, **there are significant changes in the number of conditions that have to be met in order to classify a person as materially deprived**. From a qualitative standpoint, there can be a number of expert opinions on the relevance of the presented numbers. The design intention of the indicator [1] states that the rationale for using deprivation indicators in a cross-sectional context is that current living standards and deprivation levels depend on having a command over resources and experiences over a long period, not just on the flow of income into the household this week, month, or even year. The flow of income into a household per year in regards to poverty is measured by the People at risk of poverty, after social transfers indicator with a value of 12 % for Slovakia in 2010. Taking into account the aforementioned facts, we should expect the level of material deprivation to be on par or below the level of People at risk of poverty, after social transfers. **This means cases with values above 12 % (translated into 1, 2 or 3 conditions met) can be considered as unrealistic**. On the other hand while **values that can be considered quite low** may be able to produce a good image in political terms, indicate a **limited informational value**. We consider this to be the case for 6, 7, 8 and 9 conditions met. The last two options – 11.4 % and 4.1 % (for 4 and 5 respective conditions met) are more realistic. Considering the current employment and poverty levels in Slovakia, we come to the conclusion, that four conditions are, in regards to the informational value of the indicators and the current setting of its conditions the best available options. In this setting there is a 13.3 % overlap between people in poverty and materially deprived individuals.

2.2 Changes in the conditions based on their statistical significance

Using logistic regression we will evaluate the statistical significance of the original conditions of material deprivation, as well as of conditions previously not included in the indicator. Based on the results we will analyse possible changes in the indicator values in a fashion similar to chapter 1.3. First let's take a look at the variables included in the analysis. The codes for variables we will be using are taken from EU SILC methodology [4], we also list the ratio of individuals fulfilling the individual conditions within the sample size:

Original conditions

- Inability to pay to pay rent or utility bills (HS001), 11.3 %
- Inability to keep their home adequately warm (HH050), 4 %
- Inability to pay unexpected expenses (HS060), 37.2 %
- Being unable to afford to eat meat, fish or a protein equivalent every second day (HS050), 22.4 %
- Being unable to afford a week holiday away from home (HS040), 56.4 %
- Being unable to afford a car (HS110), 17.4 %
- Being unable to afford a a washing machine (HS100), 0.6 %
- Being unable to afford a a colour TV (HS080), 0.2 %
- Being unable to afford a telephone (HS070), 0.9 %

It is interesting to observe**, that less people are unable to afford a car, compared to the ratio of people unable to fullfill their dietary requirements or being unable to pay for unexpected expences.** In fact 14,3 percent are people claiming to be unable to fullfill their dietary requirements yet able to afford a car, this applies to 25,2 % people within the sample unable to pay for unexpected expenses as well as to 41,5 % of the sample unable to afford a week holiday away from home. **Only a small amount of individuals in the sample size weren't able to afford a washing machine, a TV or a telephone, yet in all three cases around 90 % of people meeting the individual condition fullfill at least other three condition thus being classified as materially deprived.** As for the conditions not originally a part of the indicator, we picked additional dimensions of material deprivation that are included in EU SILC.

Expanded conditions

- Inability to afford seasonal clothing (SHS055), 41.3 %
- Being unable to afford a computer (HS090), 7.7 %
- Being unable to afford an internet connection (SHS095), 9.4 %
- Being unable to afford a freezer (SHS115), 4 %
- Badly lit/too dark household (HS160), 0.2 %
- Living in a noisy neighbourhood (HS170), 17.7 %
- Pollution, dirt or other environmental factors in the vicinity of the household (HS180), 20 %
- High levels of criminality and/or vandalism in the vicinity of the household (HS190), 9.7 %

In our analysis we inputted both the original and the expanded conditions into a model of logistic regression. Table 3 summarises the model, displaying only statistically significant explanatory variables for material deprivation of individuals. The model contains 9 explanatory variables. **From the original conditions**, **HS040 was removed**, despite being the one most commonly observed. In addition, **the ability to afford a car** was also **removed** due to its statistical insignificance, which perhaps may be a reflection of its anomalous overlaps we described before. **Last original condition removed was the ability to afford a colour TV**.

Taking a look at the expanded conditions, we **added the condition of inability to afford seasonal clothing** which, besides its statistical significance might be a much more important factor of material deprivation compared to owning a colour TV. **Being unable to afford an internet connection is also a statistically significant condition** when it comes to material deprivation, this condition could though be transformed into person being unable to access internet, rather than being able to afford it. **The last condition possessing a statistical significance in our model is the inability to afford a freezer**.

Our model suggests the removal of thee conditions that already were included in the indicator and an addition of other three, that weren't included. All of the logit coefficients suggest that an increase in each of the conditions will trigger an increase in the levels of the indicators overally. As for the magnitudes represented by the odds ratios, the originally included variables possess much higher magnitudes. This occurs due to the model taking into account the original structure of the material deprivation indicator.

Condition	$Logit_i$	S.E.	Wald	Sig.	$Odds_i$
HS001	5.813	.367	250.486	.000	334.552
HS050	6.137	.369	276.784	.000	462.560
HS060	6.339	.406	243.446	.000	566.323
HS070	9.689	1.520	40.639	.000	16136.43
SHS095	1.034	.117	78.079	.000	2.813
HS100	7.759	.770	101.582	.000	2342.319
HH050	5.702	.402	200.904	.000	299.596
SHS055	.608	.136	19.982	.000	1.837
SHS115	2.015	.170	139.892	.000	7.501
Constant	-14.173	.742	365.102	.000	.000

Table 3 Model of logistic regression of the material deprivation indicator

Table 4 illustrates the model fit using the Cox & Snell R Square with the value of 0.403 and the Nagelkerke R Square (generally considered to be the more suitable measure) with the value of 0.,823. **Both of these values indicate a valid model**. In the next step we will adjust the material deprivation indicator to respond to four of the nine adjusted conditions, based on the presented model.

Cox & Snell R Square	Nagelkerke R Square
.403	.823

Table 4 Model fit

The value of the adjusted material deprivation indicator for Slovakia in the year 2010 would reach 9,2 %, which is a slightly lower value compared to the original value (11.4 %). The advantage is a wider gap, when compared to People at risk of poverty, after social transfers indicator (12 %). The overlap between poverty and material deprivation in this case decreases to 11.7 %. Figure 2 illustrates the levels of material deprivation based on the number of revised conditions met. **In general we can say that the adjusted indicator of material deprivation achieves lower values in cases where less than 5 conditions have to be met**. For the rest of the cases the numbers remain almost the same. This supports using four conditions in order to validate the indicator.

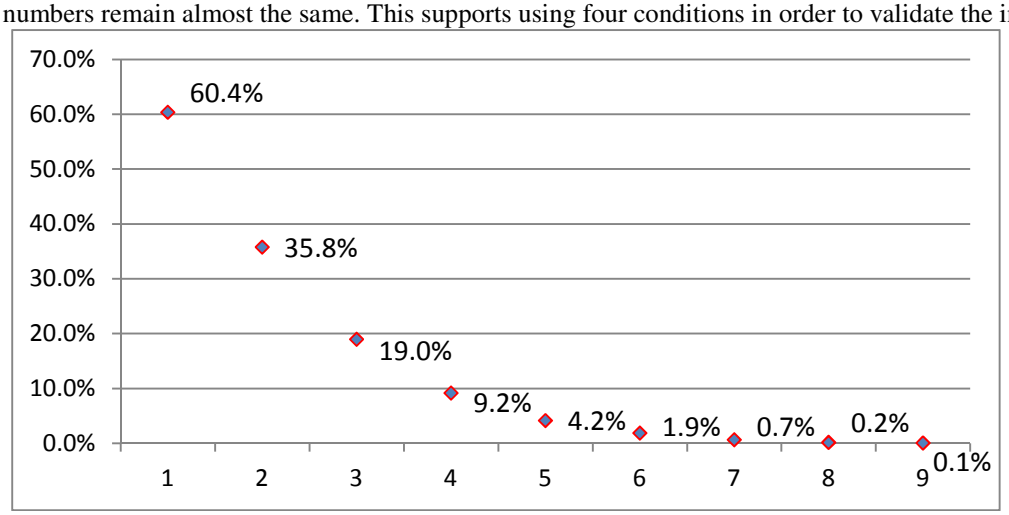


Figure 2 Levels of adjusted material deprivation based on the number of conditions which have to be met, Slovakia 2010

3 Conclusion

In this paper we conducted the analysis of the material deprivation indicator. **Based on results from a logistic regression model we constructed we proposed a removal of three conditions included in this indicator and**

suggested including three conditions that previously weren't taken into account. In Table 5 we summarize these factors.

Original condition	**Conditions after the adjustment**
Inability to pay to pay rent or utility bills	Inability to pay to pay rent or utility bills
Inability to keep their home adequately warm	Inability to keep their home adequately warm
Inability to pay unexpected expenses	Inability to pay unexpected expenses
Being unable to afford to eat meat, fish or a protein equivalent every second day	Being unable to afford to eat meat, fish or a protein equivalent every second day
Being unable to afford a week holiday away from home	Being unable to afford a telephone
Being unable to afford a car	Being unable to afford a a washing machine
Being unable to afford a a washing machine	Inability to afford seasonal clothing
Being unable to afford a a colour TV	Being unable to afford an internet connection
Being unable to afford a telephone	Being unable to afford a freezer

Table 5 Conditions of material deprivation indicator before and after the proposed

Applying the adjustment will result in a decrease of the overall value of the indicator (from 11.4 % to 9.2 %) as well as reduce the overlap between the People at risk of poverty, after social transfers indicator. **The Severely materially deprived persons indicator is also a part of an aggregation forming the values of one of the Europe 2020 headline targets** [3] **- People at risk of poverty or social exclusion**. The proposed adjustment would for People at risk of poverty or social exclusion (for Slovakia, 2010) causes a **decrease from 20.6 % to 19.2 %**. This should not be regarded as an improvement of the situation, rather as a more exact approximation of the current state. **Within our analysis we also gave evidence supporting the usage of four conditions met in order to classify the person as severely materially deprived**. **The same analysis could be conducted** for all of the other countries assuming the availability of EU SILC microdata for **all the 27 member countries** of the EU in order **to compare and generalize the recommendations for the adjustment of the indicator**.

Acknowledgements

This paper was produced with the support of the IGPM funding framework under the project number 2317122/10: Modelling of chosen indicators of Sustainable development in the context of the European Union.

References

[1] Atkinson T. et al: *SOCIAL INDICATORS: THE EU AND SOCIAL INCLUSION*, Oxford University Press, Oxford, 2002.
[2] Council of the European Union: *Renewed EU Sustainable Development Strategy Annex*, Council of the European Union, Luxembourg, 2006
[3] Council of the European Union: *Europe 2020 targets* [online], Eurostat, Luxembourg:, 2009
[4] EUROSTAT: *Sustainable development in the European Union 2011 monitoring report of the EU sustainable development strategy,* Eurostat, Luxembourg, 2011
[5] EUROSTAT: Severely materially deprived people, 2012*: Quality profile* [online], Eurostat , Brussel, 2012
[6] Haan, A., & Maxwell, **S.**: Poverty and social exclusion. *North and South IDS Bulletin* **29** (1998), 1-9.
[7] HAIR J. F. et al: *Multivariate data analysis a global perspective*, Pearson Education , New Jersey, 2010
[8] Room, G.: *Observatory on national policies to combat social exclusion* (synthesis report)., Unversity of Bath, Bath, 1990
[9] STANKOVIČOVÁ I., VOJTKOVÁ M. 2007: *Viacrozmerné štatistické metódy s aplikáciami,* Iura Edition Bratislava, 2007

Real exchange rate behavior in 4 CEE countries using different unit root tests under PPP paradigm

Ghiba Nicolae[1], Sadoveanu Diana[2], Avadanei Anamaria[3]

Abstract. This paper aims to analyze real exchange rate deflated with consumer price index under different unit root tests with and without structural breaks. We consider classical test like Augmented Dickey-Fuller and Phillips-Perron without structural break, but also unit root tests with one break and with two structural breaks, respectively Zivot-Andrews and Clemente-Montañés-Reyes techniques. Our study is based on real exchange rate analysis in 4 selected CEE countries: Czech Republic, Hungary, Poland and Romania against the Euro Zone. By many authors, real exchange rate stationarity is correlated with weak form of purchasing power parity (PPP) validity which claims that price indices between two countries/regions are convergent. The monthly data cover the 2001M01-2011M09 period. The empirical analysis provides mixed results depending on the country and the selected unit root test.

Keywords: real exchange rate, unit root tests, purchasing power parity

JEL Classification: F31, C32, E31
AMS Classification: 90C15

1 Introduction

The Purchasing Power Parity (PPP) theory is one of the most famous and the most controversial theory of exchange rate determination. PPP states that national price levels should be equal when expressed in a common currency. In most papers, the theory is tested using stationarity tests and cointegration. Nowadays, the analysis of the exchange rate sustainability is an important issue especially in view of the acceding process to the Eurozone. In order to avoid future disequilibrium, it is important for policy makers to asses the level of the national price convergence with respect to the euro market and to examine the relationship between exchange rate and prices level.

Considering the above-mentioned factors, the purpose of this study is to test the long-run validity of the Purchase Power Parity for a group of four CEE countries: the Czech Republic, Hungary, Poland and Romania. To increase the relevance of the present research and for a better understanding, we use the consumer price index (CPI) as proxy and explore several unit root tests. We structured our paper as follows: in the first section we analyze the main contributions concerning Purchasing Power Parity, in section 2 we provide a weak form description of the PPP theory, in section 3 we emphasize the methodology and in section 4 we highlight the empirical results of our study. Finally, we point out the main conclusions.

The PPP concept can be traced back to the 16th century, to the School of Salamanca, but the modern form of the theory was developed by Gustav Cassel in 1918. PPP is based on the *law of one price*, which states that once prices are converted to a common currency, the same good should sell for the same price in different countries. [10]

Based on the different econometric techniques applied over time, the empirical studies on PPP can be structured as follows: "early studies", before 1970, unit root test studies, panel data studies and cointegration-based studies.

The most representative for the early stage are the works of Frenkel [6], [7] which mainly use OLS techniques indicating the popularity of PPP theory only in countries with high inflation. The major drawbacks of the mentioned studies relate to the fact that they neglected the non-stationarity of exchange rate and prices series.

[1] "Alexandru I. Cuza" University of Iasi, Doctoral School of Economics, Romania, ghiba.nicolae@yahoo.com
[2] "Alexandru I. Cuza" University of Iasi, Dcotoral School of Economics, Romania, diana.sadoveanu@gmail.com
[3] "Alexandru I. Cuza" University of Iasi, Doctoral School of Economics, Romania, anamaria_avadanei86@yahoo.com

Later, in the second part of `80s, the notion of **stationarity** was introduced in the empirical research in order to highlight the permanent deviations from the PPP level. From this point on, the scholars had the possibility to analyze if a shock in a variable's evolution is permanent or can be absorbed over time. PPP does not hold if real exchange rate is a random walk.

The PPP approaches have been developed over time and contemporary literature is based on stationarity and cointegration using different methodologies in order to find an adequate real exchange rate determination. The long-span studies include tests on longer time periods with various currency arrangements according to international monetary system mutations. Also, the PPP analyses based on panel data and stationarity procedures are very common. One of the first PPP modern approaches of European Central-Eastern countries is the one of Choudhry's [2]. Analyzing a group of four countries (Romania, Poland, Russia and Slovenia) he found that PPP holds in its relative form only for Slovenia and Russia.

In an extensive research on a large group of countries from this region (Bulgaria, Czech Republic, Hungary, Poland, Romania and Slovak Republic in 1991-1998 period) Christev and Noorbakhsh [3] identified a long-run term relationship between prices and the exchange rate despite the law of one price, proportionality and symmetry violation.

The most recent analyses for CEE countries, which applied unit root tests are Kasman et al. [8], Telatar and Hasanov [11], Ali and Ozturk [1]. They used the traditional unit root tests and a unit root test which accounts for structural breaks. The obtained results are mixed.

For example, Kasman et al. [8] applied Lagrange multiplier (LM) test and found little evidence of PPP validity for CEE countries. The results of Telatar and Hasanov [11] research support PPP validity when considered structural changes and nonlinearities. After applying a series of unit root tests, Ali and Ozturk [1] concluded that PPP holds only for a few countries, finding weak evidence in case of other countries.

In order to test the validity of PPP for a group of 12 CEE countries and analyze the convergence of selected states towards the euro area, Kavkler et. al. [9] used a set of first generation panel unit toot tests, based on cross-sectional independency hypothesis. These tests are: Levin, Lin and Chu test, Im, Pesaran and Shin test and Fisher ADF and Fisher PP. The results of this study confirmed the stationarity of the real exchange rate, supporting the PPP theory in the long-run.

2 Purchasing Power Parity weak form description

PPP is developed having as reference the Law of One Price (LOOP) which refers to the equality between two prices of the same good, in two different countries when they are expressed in the same currency:

$$P_{i,t} = NER_t P^*_{i,t}, \qquad i = 1,2,..,n \tag{1}$$

where $P_{i,t}$ is the price of good i expressed in national currency units at moment t, $P^*_{i,t}$ the price of the same good i at moment t expressed in a foreign currency and NER_t is the nominal exchange rate at moment *t.*

If we develop the LOOP absolute form for all goods traded inside a country we obtain the PPP, also in an absolute form, as following:

$$\sum_{i=1}^{N} \alpha_i P_{i,t} = NER_t \sum_{i=1}^{N} \alpha_i P^*_{i,t} \tag{2}$$

where α_i as weights of goods included in a price index are totaling a value equal with the unity.

Integrating other influences, which include the PPP deviations, the above equation in logarithmic form can be expressed as follows (ner_t is the nominal exchange rate, p_t, p_t^* are national price indices and d_t includes deviations caused by other influence factors; all the variables are expressed in a logarithmic form):

$$ner_t = p_t - p_t^* + d_t \tag{3}$$

The deviations are associated with the real exchange rate (rer_t) volatility and the relationship becomes:

$$rer_t = ner_t - p_t + p_t^* \tag{4}$$

A major part of the empirical literature investigates the two forms of PPP theory: the strong and the weak forms. In order to analyze the weak form of the PPP stationarity tests on real exchange rates (rer_t) are applied. PPP holds in the long-run if the real exchange rate is a stationary series. The strong form is mainly investigated by applying cointegration methods and testing the proportionality and symmetry restrictions.

3 Methodology description

We admit that purchasing power parity holds in the long-run if the real exchange rate is a stationary series. A variable is stationary if it has a tendency in returning to a constant value. In other words, its trajectory must be around a mean value or around a linear trend. Economically, this means that any shock on series is temporary and it is absorbed in time. In practice, almost every variable is stationary and must be differenced. Hence, the exchange rate is nonstationary for the most cases and the series is first order integrated (requires just one differentiation). In order to test for the presence of the unit roots in the real exchange rate series were applied classical tests: Augmented Dickey-Fuller Test (ADF) and Phillips-Perron Test (PP). Also, in order to take into account the potential structural breaks in the series, which can lead to the biased results when are performed traditional tests, were applied additionally the Zivot-Andrews Test (ZA) and Clemente-Montañés-Reyes Test (CMR).

3.1 Classical unit root tests: Augmented Dickey-Fuller and Phillips-Perron

Augmented Dickey-Fuller unit root test

The econometric theory refers to a null hypothesis that claims a unit root in series. In our case, the real exchange rate is nonstationary. The most popular stationarity tests were developed by Dickey and Fuller (ADF stationarity test), respectively by Phillips and Perron [5]. The difference between them is given by the less stringent restrictions on error process for Phillips-Perron test. If the obtained t-statistisc and associated probability reflect null hypothesis acceptance, than we conclude that purchasing power parity doesn't hold.

Testing real exchange rate stationarity through Dickey-Fuller (ADF) entails three assumptions: the intercept presence, the presence of an intercept and a time trend, and finally, the absence of any deterministic element. For each supposition, we have built three different relationships, as follows: includes both a drift and a linear time trend (eq. 5), random walk with a drift (eq. 6) and pure random walk (eq. 7) :

$$\Delta rer_t = a_0 + \gamma rer_{t-1} + a_2 t + \varepsilon_t \tag{5}$$

$$\Delta rer_t = a_0 + \gamma rer_{t-1} + \varepsilon_t \tag{6}$$

$$\Delta rer_t = \gamma rer_{t-1} + \varepsilon_t \tag{7}$$

If $\gamma = 0$, than the real exchange rate sequence contains a unit root (the series is nonstationary). The test estimates a regression equation using ordinary least squares, in order to determine an estimated value for γ and associated standard error.

Augmented Dickey-Fuller is developed for p^{th}-order autoregressive process and the estimated equation is the following:

$$\Delta rer_t = a_0 + \gamma rer_{t-1} + \sum_{i=2}^{p} \beta_i \, \Delta rer_{t-i+1} + \varepsilon_t \tag{8}$$

where $\gamma = -(1 - \sum_{i=1}^{p} a_i)$ and $\beta_i = \sum_{j=1}^{p} a_j$. If $\gamma = 0$, we consider the equation being in first difference and having a unit root.

Phillips-Perron unit root test

Phillips and Perron (1988) developed ADF procedure and allowed a weaker set of assumptions regarding the error process. Also, Phillips-Perron test (PP) is powerful in rejecting the null hypothesis.

Considering the following regression equations:

$$rer_t = a_0^* + a_1^* rer_{t-1} + \mu_t \tag{9}$$

$$rer_t = \tilde{a}_0 + \tilde{a}_1 rer_{t-1} + \tilde{a}_2 \frac{t-T}{2} + \mu_t \tag{10}$$

where T is the number of observation and μ_t is the disturbance term, the PP has the followings test statistics: $Z(ta_1^*)$ for testing the hypothesis $a_1^* = 1$; $Z(t\tilde{a}_1)$ for testing the hypothesis $\tilde{a}_1 = 1$, $Z(t\tilde{a}_2)$ for testing the hypothesis $\tilde{a}_2 = 0, Z(\Phi_3)$ for testing the hypothesis $\tilde{a}_1 = 1$ and $\tilde{a}_2 = 0$.

The PP critical values are the same as in ADF stationarity test. In our analysis, we choose to use both unit root tests due to the difficulties in knowing the true data-generating process.

Zivot-Andrews unit root test with one structural break

Lately, many studies on PPP based on stationarity procedures are considering the **structural change** in order to avoid non-rejection of a unit root in that series with structural breaks, when using ADF or PP. Structural changes appear as a result of economic or financial crisis, policy changes and regime shifts.

Zivot and Andrews (ZA) [12] find a solution and identified a break point where the unit root t-statistic is the smallest. The authors tested a procedure with an estimated time of the break assuming it as an exogenous phenomenon. Therefore, they test for a unit root using three models:

- A': one-time change in the slope of the trend function is allowed:

$$\Delta rer_t = c + \alpha rer_{t-1} + \beta t + \gamma DU_t + \sum_{j=1}^{k} d_j \Delta rer_{t-j} + \varepsilon_t \quad (11)$$

- B': allows one-time change in the slope of the trend:

$$\Delta rer_t = c + \alpha rer_{t-1} + \beta t + \theta DT_t + \sum_{j=1}^{k} d_j \Delta rer_{t-j} + \varepsilon_t \quad (12)$$

- C': combines one-time change in the level and the slope of the trend function:

$$\Delta rer_t = c + \alpha rer_{t-1} + \beta t + \theta DU_t + \gamma DT_t + \sum_{j=1}^{k} d_j \Delta rer_{t-j} + \varepsilon_t \quad (13)$$

where DU_t is a dummy variable for a mean shift occurring at a possible break-date and DT_t is a corresponding trend shift variable. Also, we mention the following restrictions: $DU_t = \begin{cases} 1, if\ t > TB \\ 0, \text{otherwise} \end{cases}$ and $DT_t = \begin{cases} t - TB, \text{if } t > TB \\ 0, \text{otherwise} \end{cases}$.

The null hypothesis in all above models is α=0, which implies that y_t contains a unit root (it is non-stationary) with a drift and without any structural break. The alternative hypothesis implies α<0 and the series is considered a trend-stationary process with a one-time break at an unknown date in time. ZA test runs a regression for every possible break-date sequence and considers every point as being a potential break-date. The C model is considered as being superior to others. The main ZA unit root test advantage is that we don't need to know the exact date of a structural break.

Montanes-Clemente-Reyes unit root test with TWO structural break

Clemente Montañés and Reyes [4] developed a stationarity test, which, unlike the ZA test, takes account for two structural breaks in series. Clemente-Montañés-Reyes test estimates two models. The first model uses additive outlier and the second uses innovative outliers.

In case of an additive model, CMR test estimates equation 14 and subsequently looks for a minimum t-statistic for *α = 1 hypothesis* in the equation 15:

$$rer_t = \mu + \delta_1 DU_{1t} + \delta_2 DU_{2t} + \widetilde{rer}_t \quad (14)$$

$$\widetilde{rer}_t = \sum_{i=0}^{k} w_i DT_{b2,t-i} + \alpha \widetilde{rer_{t-1}} + \sum_{i=1}^{k} \theta_i \Delta \widetilde{rer_{t-i}} + \varepsilon_t,\ t=k+2,\ldots,T \quad (15)$$

In case of the innovative model, the estimated equation is:

$$rer_t = \mu + \delta_1 DU_{1t} + \delta_2 DU_{2t} + \vartheta_1 DT_{b1,t} + \vartheta_2 DT_{b2,t} + \alpha rer_{t-1} + \sum_{i=1}^{k} \theta_i \Delta rer_{t-i} + \varepsilon_t,\ t=k+2,\ldots,T \quad (16)$$

It also looks for a minimum t-statistic for *α=1 hypothesis.*

4 Data description and results

In this study we used data with monthly frequency covering the period from 2001M1 to 2011M9. Variables used are real exchange rates (deflated with CPI) for four CEE countries. All used variables are reflected in figure 1:

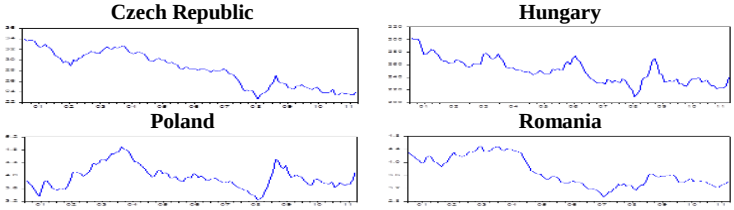


Figure 1 Real exchange rate evolution in Czech Republic, Hungary, Poland and Romania

Purchasing power parity holds in a long-run horizon if the real exchange rate deflated with different price indices is a stationary series. Null hypothesis of one unit root presence is accepted if the probability is bigger than assumed threshold of 5%. The coefficients relevance indicates the type of process: pure random walk, random walk with drift or a process with both a drift and trend time. Using ADF test results, we highlight PPP validity in long-run for Czech Republic and Hungary.

Country	**Augmented Dickey-Fuller**	
	t-statistic (Prob.)	**Conclusion (ADF)**
Czech Republic	-2.178913 (0.0288)	***Stationary in level***
Hungary	-3.919671 (0.0139)	***Stationary in level***
Poland	-2.491318 (0.1200)	Non-stationary in level
	-7.310718 (0.0000)	Stationary in first difference
Romania	-0.837202 (0.3514)	Non-stationary in level
	-7.545337 (0.0000)	Stationary in first difference

Table 1 Results of ADF unit root tests without structural breaks

Phillips-Perron is a complementary type of unit root test and it assumes a weaker set of requirement for the error process:

Null Hyphotesis: Real exchange rate has a unit root		
Country	**Phillips-Perron**	
	t-statistic (Prob.)	**Conclusion**
Czech Republic	-2.016323 (0.0423)	***Stationary in level***
Hungary	-2.702929 (0.0763)	Non-stationary in level
	-7.986771 (0.0000)	Stationary in first difference
Poland	-0.016210 (0.6755)	Non-stationary in level
	-7.396659 (0.0000)	Stationary in first difference
Romania	-1.104596 (0.2434)	Non-stationary in level
	-7.469581 (0.0000)	Stationary in first difference

Table 2 Phillips-Perron unit root test results

Using Phillips-Perron unit root test results, we find PPP validity in long-run just for Czech Republic.

Zivot and Andrews test results on the real exchange rate stationarity show PPP validity while considering structural breaks in intercept, trend or both. The null hypothesis is that the series has a unit root with structural breaks against its alternative that the series is stationary with structural breaks. ZA test is important because it offers information on that series that are non-stationary as a whole, but stationary around a break-point.

Null Hyphotesis: Real exchange rate has a unit root with a structural break in the the intercept/trend/both						
Country	**t-statistic (Prob.)**					
	Break in intercept	**Break point**	**Break in trend**	**Break-point**	**Break in both**	**Break point**
Czech Republic	-3.245670 (0.055505)	2002M10	-2.759783 (0.837176)	2003M09	Regressors are perfectly collinear	-
Hungary	***-4.750789 (0.014275)***	*2008M10*	-4.208707 (0.189965)	2008M05	***-4.726934 (0.028144)***	*2008M10*
Poland	***-3.054747 (0.048677)***	*2005M05*	-3.215168 (0.413291)	2003M02	***-4.039838 (0.012520)***	*2004M06*
Romania	***-4.342412 (0.000001)***	*2004M11*	***2.475955 (0.001648)***	*2007M12*	***--4.455401 (0.001320)***	*2004M11*

Note: Probability values are calculated from a standard t-distribution and do not take into account the breakpoint selection process;

Table 3 Zivot-Andrews unit root with a structural break test results

ZA unit root test results are as follows: PPP holds for all countries, excepting the Czech Republic when pointing a structural break in intercept and both in intercept and trend; in Romania, PPP holds for all of the assumption about the location of the structural break; PPP doesn't hold for any of the assumptions in the Czech Republic case; most of the series do have significant breakpoints, especially in the crisis period (2008 and 2008) and 2004, a period characterized by many changes in price regulation, specific to the emerging countries.

ZA unit root test results reflect different break-points in real exchange rate evolution. Using the C model, excepting the Czech Republic we find break points in all of the other countries. Structural break presence shows the specific mutations for all this emerging countries with many changes in price regulations.

Country	Additive outliers (AO)		Innovative outliers (IO)	
	t- statistic	Optimal breakpoints	t- statistic	Optimal breakpoints
Czech Republic	*-3.701*	*2005M3, 2007M10*	-4.048	*2004M10, 2007M6*
Hungary	*-4.716*	*2004M5, 2007M5*	*-4.298*	*2003M11, 2006M8*
Poland	*-4.003*	*2002M9, 2005M3*	-3.513	*2004M7, 2007M1*
Romania	*-4.898*	*2005M3, 2006M5*	*-4.564*	*2004M9, 2010M5*

** 5% critical values – two breaks: -5.49*

Table 4 Clemente-Montañés-Reyes unit root test results with double mean shift

As we could see from the table above, the Clemente-Montañés-Reyes test with two structural breaks found structural breaks for both cases: AO and IO. The period of financial crisis is the common break point for all series. Even if we identified several break points, we were unable to reject the null hypothesis of a unit root. Therefore the results of the Clemente-Montañés-Reyes unit root test with double mean shift didn't validate the PPP in long run.

5 Conclusions

Our results emphasize a growing structural breaks importance in the real exchange rate analysis using unit root tests. While ADF and PP reflect PPP validity just for one or two cases, Zivot-Andrews unit root test identifies one structural break that divide the time series in two stationary parts. Using ZA test, we highlight structural break importance and we identify their presence during the financial crisis. Contrary to these results, Clemente-Montanes-Reyes test shows that the presence of two different breakpoints leads to PPP weak form rejection.

Acknowledgements

This work was partially supported by the European Social Fund in Romania, under the responsibility of the Managing Authority for the Sectorial Operational Program for Human Resources Development 2007-2013 [grant POSDRU/CPP 107/DMI 1.5/S/78342].

References

[1] Ali, A., and Ozturk, I.: Testing Purchasing Power parity in transition Countries: Evidence from Structural Breaks, *Amfiteatru Economic Journal* **12** (2010), 190-198.

[2] Choudhry, T.: Purchasing power parity in high-inflation Eastern European countries: Evidence from Fractional and Harris-lnder cointegration tests", *Journal of Macroeconomics*, **21** (1999), 293-308.

[3] Christev, A., and Noorbakhsh, A.: Long-run purchasing power parity, prices and exchange rates in transition: The case of six Central and East European countries, *Global Finance Journal*, **11** (2000), 87-108.

[4] Clemente, J., and Montañés, A., and Reyes, M.: Testing for a Unit Root in variables with a Double Change in the Mean, *Economic Letters*, **59** (1998), 175-182.

[5] Dickey, D., and Fuller, W., A.: Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica*, **49** (1981), 1057-1072.

[6] Frenkel, J., A.: Flexible exchange rates, prices and the role of news: Lessons from the 1970s, *Journal of Political Economy*, **89** (1981), 665-705.

[7] Frenkel, J., A.: Purchasing Power Parity: Doctrinal Perspective and Evidence from the 1920s, *Journal of International Economics*, **8** (1978), 169-191.

[8] Kasman, S., and Adnan, K., and Duygu, A.: Testing the Purchasing Power Parity Hypothesis for the New Member and Candidate countries of the European Union: Evidence from Lagrange Multiplier Unit Root Tests with Structural Breaks, **46** (2010), 53-65.

[9] Kavkler, A., and Jani, B., and Darja, B.: Investigating Purchasing Power Parity in Central and Eastern European Countries: a Panel Data Approach", 2009, http://oliver.efri.hr/~euconf/2009/session2.html.

[10] Rogoff, K.: The Purchasing Power Parity Puzzle, *Journal of Economic Literature*, **34** (1996), 647-668.

[11] Telatar, E., and Hasanov, M.: Purchasing Power Parity in transition economies: evidence from the Commonwealth of Independent States, *Post-Communist Economies, Taylor and Francis Journals*, **21** (2009), 157- 173.

[12] Zivot, E., and Andrews, D.: Further evidence of great crash, the oil price shock and unit root hypothesis, *Journal of Business and Economic Statistics*, **10** (1992), 251-270.

Review of selected experiments related to the Allais paradox

Agata Gluzicka[1], Donata Kopańska-Bródka[2]

Abstract. Theorems about the rational decision making play very important role in the decision theory. According to these theorems people make their decisions by using the rule about maximum benefits. But in the literature we can find conclusions from many research and experiments which indicate that when people are making decisions, they are very often breaking that rule about maximum profits. Such research led to formulate a few paradoxes of rationality. These paradoxes emphasize the incompatibility between the theory and reality in decision making process. One of these paradoxes was presented in the 50th years in the previous century by French economist Maurice Allais.
In this article the Allais paradox is presented. With this paradox are connected among other things the certainty effect and the common consequence effect. In the paper effects mentioned above are discussed. Also the incompatibility between paradox and the expected utility theory is discussed. Moreover experiments concerning the Allais paradoxes are analyzed.

Keywords: Allais paradox, theory of expected utility, certainty effect, common consequence effect.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

One of the most famous examples of the incompatibility with the expected utility theory is experiment proposed by the French economist Maurice Allais (1953). This experiment has concerned making choices between available options in two independent situations. The original Allais paradox consists of two pairs of choices and each pair has two alternative prospects (lotteries). In the first pair of choices, one prospect is a lottery with a certain outcome and the other is risky. The risky prospect has three possible outcomes. The second pair of choices is obtained from the first by eliminating a fixed chance of winning a specific outcome. The experiment proposed by Allais is following [1, 5, 8]:

Problem 1:

A – you can received 1 000 000 dollars with probability 100%

B – you can received 5 000 000 dollars with probability 10% or received 1 000 000 dollars with probability 89% or received nothing with probability 1%.

Problem 2:

A* – you can received 1 000 000 dollars with probability 11% or received nothing with probability 89%,

B* – you can received 5 000 000 dollars with probability 10% or received nothing with probability 90%.

In this case the problem 2 is obtained from the first problem by elimination 0.89 chance of winning 1 000 000. According to the theory of the expected utility, if A is preferred to B, then A* should be preferred to B*. However, in the problem 1 most people chosen the option A and in the second problem the option B* is preferred. Maurice Allais proved that the choosing the option A and B* reveals preferences which are inconsistent with the utility theory axioms. Let $u()$ denote the utility function and $u(0)=0$, then the incompatibility can be shown in the following way:

- because A is preferred to B then

[1] University of Economics in Katowice, Department of Operations Research, ul. 1 Maja 50, 40-287 Katowice, agata.gluzicka@ue.katowice.pl.
[2] University of Economics in Katowice, Department of Operations Research, ul. 1 Maja 50, 40-287 Katowice, donata.kopanska-brodka@ue.katowice.pl.

$$u(1\,000\,000) > 0.1u(5\,000\,000) + 0.01u(0) + 0.89u(1\,000\,000) \Leftrightarrow 0.11u(1\,000\,000) > 0.1u(5\,000\,000)$$

- because B* is preferred to A* then

$$0.1u(5\,000\,000) + 0.9u(0) > 0.11u(1\,000\,000) + 0{,}89u(0) \Leftrightarrow 0.1u(5\,000\,000) > 0.11u(1\,000\,000)$$

We received two contrary inequalities, what proves incompatibility between expected utility theory and paradox. The Allais paradox is usually interpreted as evidence against the "sure thing" principle or independence axiom of expected utility theory.

2 Allais certainty effect and the common consequence effect

In the original Allais paradox one of the lotteries should be a lottery with a certain outcomes. Many research concerned similar experiments. Such experiments were also analyzed by Kahneman and Tversky in the paper: "Prospect theory: An Analysis of Decision under Risk" [8]. This paper concerned the critics of the expected utility theory. Authors were emphasized that choice risky prospects conduct to some effects which are incompatible with the general rules of expected utility theory. In the expected utility theory, the utilities of outcomes are weighted by probabilities of their occurrences. In particular, people overweight certain outcomes to outcomes which are only probable. This tendency is called certainty effect or Allais certainty effect and it is contributing to that the risk aversion appears in the choice with the certain outcomes and the risky attitude appears in the choice connected with some losses. They considered following problems [8]:

Problem 3: Choose between

A: 2500 with probability 0.33
2400 with probability 0.66
0 with probability 0.01

B: 2400 with certainty

Problem 4: Choose between

A*: 2500 with probability 0.33
0 with probability 0.67

B*: 2400 with probability 0.34
0 with probability 0.66

In this experiment most respondents (82%) chose the lottery B in the problem 3, but in the problem 4 the lottery A* was preferred to the lottery B*. The lottery A* was chosen by the 83% of respondents. This example evidently violates **t**he rules of the expected utility theory.

Experiment presented above is also an example of the common consequence effect, because the problem 4 can be obtained from the problem 3 by elimination a 66% of chance to getting the value 2400 in both lotteries. This reduction changes the lottery with the certain outcomes to the lottery with the probable outcomes.

Among many experiments concerning the Allais paradox, we can find such as which don't prove incompatibility with theory of expected utility. Some examples of such experiments were proposed by Shu Li [10]:

Problem 5: Choose between

A: You can have 3000 with certainty

B: You have 10% chance of getting 3000, 45% chance of getting 6000, and 45% chance of getting nothing.

Problem 6: Choose between

A*: You have 90% chance of getting 3000, and 10% chance of getting nothing.

B*: You have 45% chance of getting 6000, but 55% chance of getting nothing.

In the problem 5 we have lottery with the certain outcome but in this case author didn't receive the violation of rules of the expected utility theory. In this experiment, over 63% of respondents chose the lottery A and 89% chose the lottery A*.

The next experiment is another example of negation the conclusions formulated by Kahneman and Tversky. Experiment provided by Shu Li indicated that sometimes the risky lottery is preferred to the lottery with certain outcome [10]:

Problem 7: Choose between

A: You can have 150 with certainty

B: You have 11% chance of getting 150, 24% chance of getting 230 but 55% chance of

getting 120.

Problem 8: Choose between

A*: You have 89% chance of getting 150 and 11% chance of getting nothing.

B*: You have a 24% chance of getting 230, but a 65% chance of getting 120 and 11%chance of getting nothing.

In the above example we don't have the certainty effect because 65% of respondents chose lottery B. In the problem 8 lottery B* was preferred (62% of respondents chose this lottery) and that indicated that in this case respondents chose lotteries also without violation the expected utility theory. It means most people preferred the risky lottery so the winning of the 150 is less attractive than random lottery of roughly equal expected value. Even though a certain result was reduced to a probable one, subjects generally chose the risky option.

The other types of the Allais paradox are lotteries without the certain outcome. In this case we are considering two pairs of lotteries where the one pair is obtained from the other by eliminating some chances of winning. Let's consider the following example [10]:

Problem 9: Choose between

A: You have 21% chance of getting 3000, 78% chance of getting 2500 but 1% chance of

getting nothing

B: You have 21% chance of getting 3000 and 79% chance of getting 2400.

Problem 10: Choose between

A*: You have 78% chance of getting 2500 and 22% chance of getting nothing.

B*: You have 79% chance of getting 2400 and 21% chance of getting nothing.

This example is some type of answer to the statement formulated by Kahneman and Tversky: a reduction of the probability of an outcome by a constant factor had more impact when the outcome was initially certain than when it was merely probable. In this case preferred lotteries were: B (66%) and A* (69%), what denied the above statement.

3 Common ratio effect and the reverse common ratio effect

With the Allais paradox is also connected property called the common ratio effect. In this case the one problem is obtained from the other by reducing probability by the same ratio in both lotteries. Few examples of the common ratio effect we can find in the article of Kahneman and Tversky who analyzed these following lotteries [8]:

Problem 11: Choose between

A: 4000 with probability 0.80	B: 3000 with certainty
0 with probability 0.20	

Problem 12: Choose between

A*: 4000 with probability 0.20	B*: 3000 with probability 0.25
0 with probability 0.80	0 with probability 0.75

Similarly as in the most analyzed problems, more than half of respondents had violated rules of the expected utility theory. In the problem 11 most respondents (80%) chose the lottery B and in the problem 12 about 65% respondents lottery A* preferred to the lottery B*.

Let's note that the lottery A* we can express in the following way: (A, 0.25). The substitution axiom of the expected utility theory asserts that if lottery B is preferred to the lottery A then any probable combination lottery (B, p) must be preferred to the combination (A, p). But in the above example most people didn't apply this rule. Authors noted that reduction of the probability from 1 to 0.25 have greater effect than the reduction of the probability from 0.8 to 0.2. In this case the reduction of probabilities additionally was changed the lottery with the certain result to lottery with the probable result.

An example of the common ratio effect is presented also in the next problems [8]. In this case problem 14 is constructed from the problem 13 by reduction probabilities by the same coefficient equal 450.

Problem 13: Choose between

A: 6000 with probability 0.45	B: 3000 with probability 0.90
0 with probability 0.55	0 with probability 0.10

Problem 14: Choose between

A*: 6000 with probability 0.001	B*: 3000 with probability 0.002
0 with probability 0.999	0 with probability 0.998

Let's note that in the problem 13 the probabilities of winning are significant and they are equal 0.90 and 0.45 respectively. In this case most respondents (86%) chose the lottery B what means they preferred that lottery where the winning is more probable. But in the problem 14, when we have only a small chance of winning (only 0.001 and 0.002 respectively), most people chose this lottery which offers bigger amount to win (the lottery A* was preferred by the 73% of respondents).

The problem of the common ratio effect was also analyzed by Blavatsky. He has expanded this problem and he started distinguish common ratio effect from the reverse common ratio effect [3]. He analyzed lotteries presented in the table 1.

Experiment	A	B	A*	B*
1	($60, 1)	($100, 0.75)	($60, 0.33)	($100, 0.25)
2	($50, 1)	($100, 0.75)	($50, 0.33)	($100, 0.25)
3	($40, 1)	($100, 0.75)	($40, 0.33)	($100, 0.25)
4	($30, 1)	($100, 0.25)	($30, 0.33)	($100, 0.08)
5	($20, 1)	($100, 0.25)	($20, 0.33)	($100, 0.08)
6	($10, 1)	($100, 0.25)	($10, 0.33)	($100, 0.08)

Table 1 Pairs of decision problem used in experiments in [3]

All pairs of lotteries considered in these experiments were similar. Lottery A was a sure payoff lottery, lottery B was a risky, lottery A* was safer lottery and lottery B* was riskier lottery. Pairs A and B Blavatsky called scaled up lotteries and pairs A* and B* scaled down lotteries. Lotteries A* and B* were obtained from the lotter-

ies A and B respectively by reduction of the probabilities of winning (in every problem the probability of the winning is divided by 3). In every case when the preferred lotteries were: lotteries A and B* or lotteries A* and B, it was the violation the expected utility theory. The first pair of lotteries is an example of the common ration effect and the second pair of the lotteries is an example of the reverse common ration effect. Blavatsky indicated that the reverse common ratio effect occurred more often than the classical common ratio effect. The similar experiment had been conducted for different amount of winning, but obtained results only confirmed conclusions received before.

4 Lotteries with the real payoffs and few other ways to reduce violation of the utility theory

Most research about the Allais paradox concerned the hypothetical lotteries. But in some case respondents could receive the real payoffs. The amounts of payoffs were adequate to outcomes of lotteries preferred by respondents. If analyzed lotteries were hypothetical (without money) then the percent of the violation was higher than the violation in the case when the respondents received real payoffs. Such research was presented in the article of Harrison, which analyzed the following lotteries [6]:

Problem 15: Choose between

A: You can have a $5 with certainty

B: You have a 0.01 chance of getting nothing, a 0.89 chance of getting $5% and 0.10 chance of getting $20.

Problem 16: Choose between

A*: You have 0.89 chance of getting nothing and 0.11 chance of getting $5

B*: You have 0.90 chance of getting nothing and 0.10 chance of getting $20

The choice of lotteries B and B* is compatible with the rules of the expected utility theory. Such decision was made by the 65% of respondents. But the choice of lotteries A and B* is a violation of these rules and such decision was made by the 35%. These results were received in experiment without money. In the case when the participants knew that their decisions were connected with the real money, 85% of respondents make his decision according to the rule of maximizing the expected utility and the percentage of violation this rule decreased to the 15%.

The similar experiments, where the results for the hypothetical lotteries were compared with the results with the lotteries with real payoff were presented in [4]. They analyzed a few lotteries with values of the winning equal $5 and $10 and with different chance of winning. One of the analyzed experiments was following:

Problem 17: Choose between

A: You can have a $5 with certainty

B: You have a 0.20 chance of getting nothing, a 0.80 chance of getting $10

Problem 18: Choose between

A*: You have a 0.75 chance of getting nothing and 0.25 chance of getting $5

B*: You have a 0.80 chance of getting nothing and a 0.20 chance of getting $10

In this case for the lotteries without money, the percent of the violation was equal 36% and when respondents knew about the real payoffs only 8% respondents violated the rule of the expected utility theory. The received conclusion was that the monetary incentives can have a systematic effect on the lottery choice. Specifically, violation of expected utility theory was significantly reduced when lotteries were real rather than hypothetical.

The value of the winning is a one of factors which influence to the result of experiment. Examples of such experiments are presented by the Huck and Muler [7]. These authors analyzed lotteries with the original values proposed by the Maurice Allais ($1 000 000 and $5 000 000) and also lotteries with the lower amounts equal $5 and $25. Additionally, they analyzed the example of lotteries with the real payoffs. In the case of lotteries with the original values of winning, the percent of the violation was equal 30%. When the decision was connected

with the lower values only 15% respondents took decision inconsistent with the rule of the maximal utility. This and the other similar research indicated that if payoff is closer to the reality (for example payoff is close to the average level of income) then we receive the higher level of violation of the expected utility theory.

Many other factors influence to the level of violation the theory of expected utility. The other research was connected for example with the form of the presented problems. Other results we receive in the case when we describe the possible lottery (like in this paper) and the other results we receive when lottery are presented in graphical form. The other form was proposed by Conlisk [5], which presented the three-step Allais paradox. By using this form of the paradox, author decreased the level of violation from the 50% (for the Allais paradox in classical form) to 28%.

Also, very important is adequate matching probabilities to the amount which we can win. Some examples of experiment connected with the probabilities in the Allais paradox were presented in [2]. The results of experiment are also dependent from the respondents. The Allais paradox can be analyzed in group of people connected in some way with the decision theory as well as in the group of people which hadn't any common with this field. Most experiments were conducted in two or more groups, but the difference between obtained results in particular groups wasn't significant [3, 10]. About the Allais paradox we can also talk in the case of lotteries without the monetary outcomes. We can find some examples about mountaineering expedition problem [9] or lotteries where you can win tour [8].

5 Conclusion

Empirical research has shown that people often violate the implications of expected utility when choosing between risky alternatives. A few paradoxes of rationality had been formulated on the basis of these observations. Allais paradox is one of them. Incompatibility of this paradox with the axiomatic of utility theory had given motivation to the development of alternative theories such as original prospect theory [8], rank dependent utility [9], cumulative prospect theory [11, 12] and weight models such as TAX and RAM [2]. In all of these theories researchers try to explain the Allais paradox.

References

[1] Andreoni J., Sprenger C.: Certain and uncertain utility the Allais paradox and five decision theory phenomena. *Levine's Working Paper Archive*, (2010), 1-25

[2] Birnbaum M.H., Schmidt U.: Allais paradoxes can be reversed by presenting choices in canonical split form. *Kiel Working Paper* **1615** (2010), 1-28

[3] Blavatsky P.R.: Reverse common ratio effect. *Journal of Risk Uncertain* **40** (2010), 219-241

[4] Burke M.S., Carter J.R., Gomniak R.D., Ohl D.F.: An experimental note on the Allais paradox and monetary incomes. *Empirical Economics* **21** (1996), 617-632

[5] Conlisk J.: Three Variants on the Allais Experiments. *The American Economic Review* **79** (1989), 392-407

[6] Harrison G.W.: Expected utility and the experimentalists. *Empirical Economics* **19** (1994), 223-253

[7] Huck S., Muller W.: Allais for all: revisiting the paradox. *ELSE Working Papers* **289** (2007), 1-22

[8] Kahneman D., Tversky A.: Prospect theory: An Analysis of Decision under Risk. *Econometrica* **47** (1979), 263-292

[9] Quiggin J.: A theory of anticipated utility. *Journal of Economic Behavior and Organization* **3** (1981), 323-343

[10] Shu L.: What is wrong with Allais' certainty effect? *Journal of Behavioral Decision Making* **6** (1993), 271-281

[11] Tversky A., Kahneman D.: Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* **5** (1992), 297-323

[12] Wakker P, Tversky A.: An axiomatization of cumulative prospect theory. *Journal of Risk and Uncertainty* **7** (1993), 147-175

Business cycle correlation of the CEEC and the Euro area: some methodological controversy

Ladislava Grochová [1], Petr Rozmahel [2]

Abstract. The paper focuses on some methodological features of the business cycle correlation measuring. Particularly, a spurious influence of detrending techniques used when analyzing the classical and growth business cycles is examined in the paper. The results give some evidence of different impact of detrending filters on final measures of correlation since uneven volatility, autocorrelation and other characteristics in detrended time series are produced by the filtering techniques. This might lead to a biased or distorted interpretation of the results when assessing business cycle correlation.

Keywords: business cycle, correlation, detrending, Hodrick-Prescott filter, band-pass filter,

JEL Classification: C18, C82, E32
AMS Classification: 91G70

1 Introduction

During past two decades a vast variety of research papers and analytical studies on business cycle similarity measuring were published in economic literature. The business cycle similarity and convergence was assessed mainly with regards to analyzing the economic and monetary integration processes in Europe. Particularly, business cycles of candidate countries aceeding the Euro zone were analyzed and compared with the member countries' cycles. The characteristic of business cycle similarity and shock asymmetry were defined in a frame on the "New" Optimum Currency Areas Theory [10]. Technically, the classical and growth business cycles have to be identified first in order to apply some measure of similarity. Burns and Mitchell [2] define the classical business cycles as cyclical fluctuations with decline and growth phases in an absolute level of aggregate economic activity of a nation. The growth cycles are considered as fluctuations of cyclical component of analyzed macroeconomic time series around its trend [9]. Accordingly, a selected detrending technique to dissect the cycles and trends in the data time series need to be applied.

The OCA literature provides a general theoretical framework for business cycle similarity assessment [11]. However, the theory does not give a clear recommendation of a concrete method or technique to be applied. Thus a vast variety of detrending and cycle identification method was used in the past literature. Fidrmuc and Korhonen [7] used a meta-analysis of the business cycle correlation between the Euro area and Central and Eastern European countries (CEEC) analyzing 35 selected publications on business cycle synchronization in Europe. In the analysis they compare the datasets, methods including detrending techniques and results of chosen topical papers. Among others they found more conservative and careful estimation results in studies by the central banks comparing to those by research institutions and academia. There is also a lot of econometric and statistical literature dealing with impact of using methods such as filters on resulted time series including business cycles (e.g. [1], [4], [5], [13], [12], [6], and [14])

In our paper we concentrate on the role of detrending techniques and filters in the business cycle correlation analysis. Particularly, we focus on evaluation of spurious impacts of various detrending techniques on macroeconomic time series when estimating the business cycles correlation. The paper is structured as follows. After the introduction the second session describes the data and used methodology. The correlation of business cycles and detrended time series characteristics are compared in the third section. The fourth section concludes.

2 Data and methodology

The monthly data of industrial production in 1994-2011 sourced from IFS IMF is used when analyzing the business cycles. Germany, France, Italy, Spain and Portugal were selected as the representatives of core and peripheral Euro area countries. Czech Republic, Hungary, Poland and Slovakia complete the sample of CEE countries.

[1] Mendel University in Brno, department of Economics, Zemědělská 1, 613 00 Brno, Czech Republic, e-mail: ladislava.grochova[at]mendelu.cz.
[2] Mendel University in Brno, department of Economics, Zemědělská 1, 613 00 Brno, Czech Republic, e-mail: petr.rozmahel[at]mendelu.cz.

The Hodrick-Prescott filter, Baxter-King band pass filter using for identification of growth cycles and first differencing of logarithms for classical cycles were applied in the analysis.

There is a lack of consensus in economic literature for the classical business cycles identification. The first differencing method is used in studies on classical as well as growth cycle. As Canova [5] points out the first differencing of logarithms is an appropriate method producing stationary time series provided that the trend component represents the random walk process, cyclical component is stationary and both components are not correlated. In addition the unit root of y_t is supposed due to systematic component y_{t-1}, for

$$y_t = y_{t-1} + \varepsilon_t \tag{1}$$

For which a trend is defined as $g_t = y_{t-1}$ and $\hat{c}_t = y_t - y_{t-1}$.

The Hoddrick-Prescott filter [8] enables to dissect the stochastic trend and uncorrelated cyclical component in analyzed time series solving the minimum problem

$$\min_{\{g_t\}_{t=1}^{T}} \sum_{t=1}^{T} c_t^2 + \lambda \sum_{t=1}^{T} \left[(g_{t+1} - g_t) - (g_t - g_{t-1}) \right]^2, \ \lambda > 0 \tag{2}$$

We apply the smoothing parameter λ=14400 (monthly data) penalizing the variability in the growth component. The value of smoothing parameter λ can be adjusted according to the business cycle definition. The greater λ the smoother trend identified in the time series.

Finally the frequency domain band pass filter suggested by Baxter and King [1] was applied in the paper. The assumption drawn from Burns-Mitchell [2] of the suggested filter is the cycle no shorter than 1.5 year and no longer than 8 years. The Baxter-King filter is basically the two-sided symmetric linear filter in a form of a combination of low-pass ad high pass filter passing by the components with periodical fluctuations between 6-32 quarters of the spectrum. The components with fluctuations of higher or lower frequencies are removed by the filter. The authors define the ideal truncation period K=12 for quarterly data and K=36 for monthly data. This implies eliminating of 12 and 36 observations, respectively, at the beginning and the end of the time series [3]. The length of the truncation period improves the equality of resulting time series decomposition but bigger loss of data.

Concerning the methods used the cross-correlation analysis of analyzed countries is applied first. The resultant correlation matrices are compared in case of all three detrending techniques applied in order to point out significant differences in resultant correlation coefficients when using various filters. Next, the business cycle of selected country identified with selected detrending techniques is described and characterized. Finally, the resultant detrended time series approximating the business cycles are analyzed and compared focusing on filters' properties such as autocorrelation, standard deviation and a frequency of turning points in order to examine various effects of selected detrending techniques.

3 Business cycle correlation of CEEC and the Euro area

3.1 Cross correlation

Table 1 includes the cross correlation coefficients with p-values of all selected countries. The coefficients resulted from using the Hedrick-Prescott (HP) and Baxter-King band pass filter (BK) are compared in a cross correlation matrix. Table 2 completes the results obtained with applying the first differencing technique.

	bk_fr	bk_it	bk_de	bk_es	bk_pl	bk_hu	bk_cz	bk_sk	
hp_fr	1	0.7645	0.6945	0.6865	0.511	0.7227	0.5081	0.6671	bk_fr
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	
hp_it	0.9054	1	0.8091	0.7133	0.4939	0.7865	0.5733	0.5899	bk_it
	(0.0000)		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	
hp_de	0.9204	0.909	1	0.6265	0.4208	0.8075	0.5488	0.6043	bk_de
	(0.0000)	(0.0000)		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	
hp_es	0.8689	0.8738	0.8586	1	0.6061	0.6917	0.4653	0.6401	bk_es

	(0.0000)	(0.0000)	(0.0000)		(0.0000)	(0.0000)	(0.0000)	(0.0000)	
hp_pl	0.6087	0.6405	0.5925	0.665	1	0.4694	0.6198	0.4862	bk_pl
	(0.0000)	(0.0000)	(0.0000)	(0.0000)		(0.0000)	(0.0000)	(0.0000)	
hp_hu	0.8599	0.8353	0.8816	0.8416	0.5826	1	0.4436	0.5079	bk_hu
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)		(0.0000)	(0.0000)	
hp_cz	0.7563	0.7853	0.7728	0.7095	0.6272	0.6805	1	0.4982	bk_cz
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)		(0.0000)	
hp_sk	0.6258	0.7121	0.6319	0.6229	0.5917	0.5394	0.682	1	bk_sk
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)		
	hp_fr	hp_it	hp_de	hp_es	hp_pl	hp_hu	hp_cz	hp_sk	

Table 1 Statistical characteristics of industrial production cycles identified with Hodrick-Prescott and Baxter-King band pass filter; source: author's calculations; *p*-values in parentheses,

	fd_fr	fd_it	fd_de	fd_es	fd_pl	fd_hu	fd_cz	fd_sk
fd_fr	1							
fd_it	0.4480	1						
	(0.0000)							
fd_de	0.3534	0.3148	1					
	(0.0000)	(0.0000)						
fd_es	0.3121	0.3621	0.1949	1				
	(0.0000)	(0.0000)	(0,0031)					
fd_pl	0.2641	0.2108	0.0835	0.2130	1			
	(0.0001)	(0.0013)	(0.2080)	(0.0012)				
fd_hu	0.2121	0.1573	0.2235	0.1784	0.1136	1		
	(0.0013)	(0.0174)	(0.0007)	(0.0069)	(0.0870)			
fd_cz	0.2077	0.2343	0.1214	0.1799	0.3961	0.2167	1	
	(0.0017)	(0.0004)	(0.0672)	(0.0065)	(0.0000)	(0.0010)		
fd_sk	0.2153	0.2067	0.1337	0.1570	0.4530	0.1254	0.4063	1
	(0.0011)	(0.0017)	(0.0437)	(0.0177)	(0.0000)	(0.0588)	(0.0000)	

Table 2 Statistical characteristics of industrial production cycles identified with first differencing; source: author's calculations; *p*-values in parentheses.

Comparing the results in both tables one might conclude that the Euro zone member countries reveal higher cross correlation than the CEE countries. Using he HP filter all countries seem to share a kind of European business cycle since all correlation coefficients higher than 0.5 with high statistical significance. Even the CEE countries such as the Czech Republic and Hungary show business cycles highly correlated to Germany as a benchmark for the Euro zone average. Results in case of using the BK filter are very similar to those of HP filter. Apart from few exceptions the cross correlation confirms highly similar cycles in the whole analyzed period of 1994-2011. However, the situation is dramatically different in case of applying the first differencing technique. All cross correlation coefficients are significantly lower than those of using HP and BK filters. The resultant coefficients even give an evidence of uncorrelated cycles with high statistical significance. Let's recall that first differencing of logarithm of the macroeconomic data implies an approximation of the growth rates, which is commonly used in studies on economic integration. Since the underlying OCA theory does not provide a clear methodological recommendation on a concrete business cycle identification method, the researchers differ in choosing the detrending techniques, which might influence the final correlation results. Accordingly, the final interpretation of numerical results such as business cycle correlation might be spurious without consideration of different methodological properties.

3.2 Detrending techniques influence

To shed some light on the rationale of a significantly different correlation results obtained when using various filters we now focus on selected statistical properties of applied detrending techniques. The Figure 1 shows the detrended time series when using first differencing, Hodrick-Prescott and Baxter-King filters.

The first differencing technique produces cycles of lower volatile time series measured with standard deviation and more frequent turning points. HP and BK filters generate rather smoother cycles. This is more obvious when using quarterly data of GDP which has less strength in the higher frequencies of the spectra. In our paper we use monthly data of industrial production, which produce frequent turning points in the detrended time series.

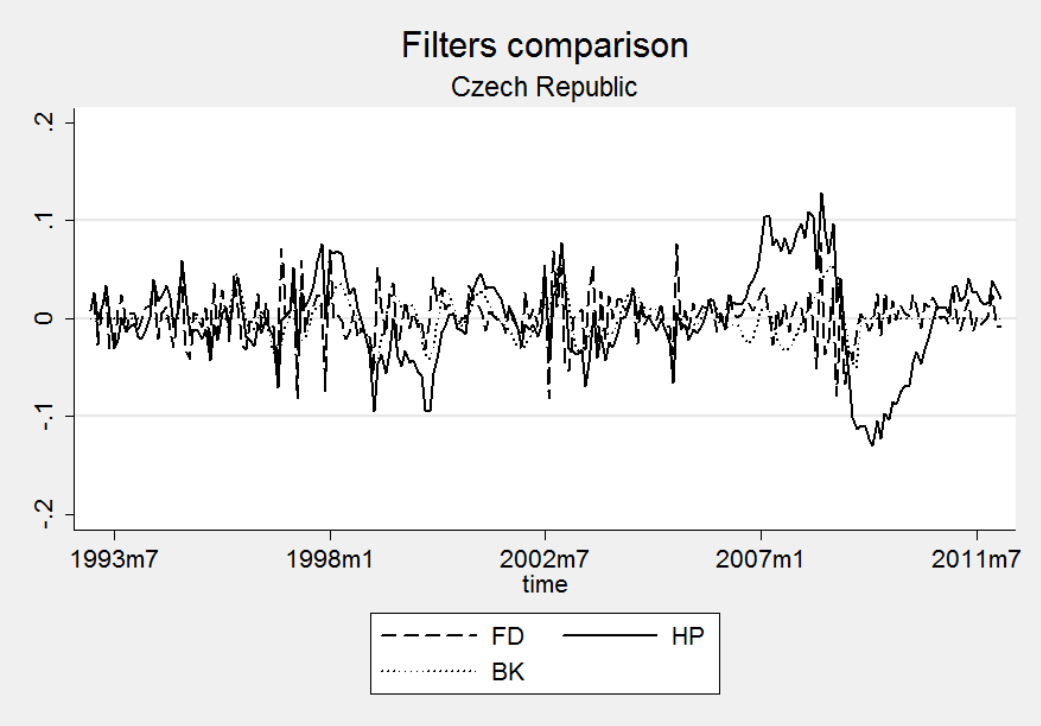


Figure 1 First differencing, Hodrick-Prescott and Baxter-King filters comparison: the case of Czech industrial production cycles

Variable	Obs	Mean	Std. Dev.	Min	Max	r(1)	r(2)	r(3)	Turning Points
hp_fr	157	0.003586	0.02298	-0.08286	0.07448	0.803804	0.749725	0.666813	26
hp_it	157	0.005904	0.032947	-0.0882	0.108008	0.880411	0.811609	0.747069	21
hp_de	157	0.00698	0.036678	-0.12421	0.103172	0.869236	0.802636	0.752651	22
hp_es	157	0.004852	0.034588	-0.0835	0.105498	0.885731	0.827801	0.774792	13
hp_pl	157	0.005332	0.042725	-0.09055	0.13098	0.546223	0.598242	0.638061	41
hp_hu	157	0.005939	0.050413	-0.12889	0.131484	0.854477	0.774666	0.72127	22
hp_cz	157	0.006613	0.045914	-0.11313	0.127965	0.784187	0.703349	0.59827	27
hp_sk	157	0.005796	0.055276	-0.16623	0.179847	0.742889	0.669494	0.58775	33
fd_fr	157	-3.6E-05	0.012765	-0.04502	0.037134	-0.31887	0.177753	0.094873	90
fd_it	157	-0.00095	0.014455	-0.04922	0.038412	-0.13731	0.14536	0.225021	88
fd_de	157	0.000901	0.015565	-0.07873	0.03011	-0.07068	0.083816	0.254739	89
fd_es	157	0.000213	0.01552	-0.04134	0.063776	-0.23481	0.107721	0.205314	91
fd_pl	157	0.004727	0.040152	-0.13158	0.12687	-0.56217	0.0071	0.3808	109
fd_hu	157	0.004749	0.025346	-0.12142	0.064458	-0.23306	0.080036	0.040149	81

fd_cz	157	0.002482	0.028886	-0.08162	0.077813	-0.31769	0.106423	0.026169	75
fd_sk	157	0.003056	0.037486	-0.16809	0.125571	-0.40684	0.086709	0.138359	98
bk_fr	157	0.000648	0.009455	-0.01988	0.036833	0.900354	0.698133	0.485087	27
bk_it	157	0.000282	0.014976	-0.04358	0.05082	0.885511	0.682071	0.460684	17
bk_de	157	0.001103	0.015824	-0.05534	0.074079	0.862272	0.632629	0.401136	19
bk_es	157	-0.0005	0.011101	-0.04244	0.037015	0.858594	0.590537	0.314767	26
bk_pl	157	-0.00032	0.018296	-0.05226	0.049488	0.874419	0.593139	0.286072	21
bk_hu	157	0.000438	0.01985	-0.07291	0.070847	0.853719	0.580257	0.302516	24
bk_cz	157	0.000475	0.022351	-0.04989	0.055034	0.852671	0.530043	0.178718	23
bk_sk	157	0.000388	0.027297	-0.08906	0.094131	0.865441	0.578016	0.25625	19

Table 3 Statistical characteristics of industrial production cycles identified with the chosen techniques, source: author's calculations

The table 3 summarizes selected properties of detrended time series of monthly industrial production in 1996-2008. The analyzed time series is six years shorter since the Baxter-King band pass filter needs three years (truncation period) to be cut off at the beginning and the end of the analysed time period. The analyzed properties include selected descriptive statistics, autocorrelation measures and also number of turning points at the detrended time series. The monthly data of industrial production does not allow detecting the lowest standard deviation in case of first differencing as suggested in Baxter-King [1]. The measures of autocorrelation and number of turning points provide evidence of different properties of selected filters. The Hodrick-Prescott and Baxter-King band pass filter reveal remarkably higher correlation than first differencing technique. In addition, the number of turning points is 3-5 times higher when using first differencing then HP and BK filters. On the contrary HP and BK filters produce time series with very comparable number of turning points.

As suggested in Baxter-King [1] first differencing emphasises high frequencies and down weights low frequencies of the time series spectra. This implies lower correlation and autocorrelation in detrended time series when using first differencing. HP and BK filters work as band pass filters. HP filter acts as a high-pass filter which leaves components of higher frequencies in the time series whereas the BK-BP removes them. The correlation and autocorrelation measures are higher when using quarterly and monthly data of GDP and other aggregate activity indicators that do not have much of high frequencies in the spectra. Also volatility measured with standard deviation is higher than in case of first differencing in case of quarterly data. Using monthly data of industrial production our results confirm those of Baxter-King [1]. Standard deviation of cycles when using first differencing is not significantly lower than of the other filters. This might be the impact of using monthly data since Baxter-King [1] also found comparable standard deviations in time series when using monthly data of inflation rate and other selected indicators.

4 Conclusion

In our paper we intended to provide some evidence of an influence of selected detrending techniques on business cycle correlation measuring. Three selected detrending techniques were applied to identify cycles of monthly industrial production in selected EU countries including the CEEC. The results of business cycles cross correlation show significant differences when using band pass filters and first differencing. Whereas business cycle correlation identified with using Hodrick-Prescott and Baxter-King band pas filters are highly correlated, cycles detrended with first differencing are uncorrelated. This might imply spurious interpretation when assessing the business cycle similarity from OCA theory perspective. Analysis of selected statistical properties of filtered time series show significant differences in measures of autocorrelation and number of turning points. First differencing produce time series with low autocorrelation and highly frequented turning points. Band pass filters produce rather smoother cycles in terms of significantly lower number of turning points and higher autocorrelation measures. All three filters emphasize and down weight dissimilar frequencies of the time series spectra. The results also indicate potentially different properties of filtering techniques when using quarterly and monthly data. Since there is a lack of consensus on business cycle identification technique in the literature, the assessment of business cycle similarity might lead to inappropriate interpretation due to uneven impact of chosen filters. Accordingly, our recommendations are aimed at careful interpretation of results with respect to applied methodology.

References

[1] Baxter, M., and King, R. G.: Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series. *The Review of Economics and Statistics* **4** (1999), 575–593.

[2] Burns, A.F., and Mitchell, W.C.: Measuring Business Cycles: Vol. 2 of Studies in Business Cycles. New York: NBER, 1946.

[3] Buss, G.: Preliminary Results on Asymmetris Baxter-King Filter. *Journal of Applied Mathematics* 4 (2011), 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 Reserev Bank of New York Staff Report 289 (2007), 34 p.

[7] Fidrmuc, J., and Korhonen, I.: Meta-Analysis of the business cycle correlation between the Euro area and CEECs. *Journal of Comparative Economics* **34** (2006), 518–537.

[8] Hodrick, R., and Prescott, E.C.: Post-War U.S. Business Cycles. Pittsburgh PA: Carnegie Mellon University, Working Paper, 1980.

[9] 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.

[10] Mongelli, P. F.: New Views on The Optimum Currency Area Theory: What is EMU Telling Us? Working Paper ECB 138, 2002.

[11] Mundell, R. A.: A theory of Optimum Currency Areas. *The American Economic Review* **51** (1961), 657–665.

[12] 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.

[13] Stock, J. H., Watson, M. W.: Business Cycles Fluctuations in U.S. Macroeconomic Time Series. In.: *Handbook on Macroeconomics* (Taylor J. B., and Woodford, M., eds.). Elsevier, Amsterdam, 1999, 3–64.

[14] Watson, M.: How Accurate Are Real-Time Estimates of Output Trends and Gaps? *Economic Quarterly* **93** (2007), 143–161.

Predicting bankruptcy of companies based on the production function parameters

David Hampel[1], Jan Vavřina[2], Jitka Janová[3]

Abstract. In this paper we want to determine economic performance of the sample of companies using four-parametric production function. The available microeconomic data set was obtained by the secondary research utilising database Amadeus and financial statements issued in the Business Register of the Czech Republic. We focus on business entities within the agribusiness industry sector that is in comparison with other sectors considered to be the economically weakest one. The initial part of the work deals with identification of proper economic indicators of respective business entities which can be used as input factors of production function. To the estimation and testing of production function parameters is necessary to use nonlinear techniques, including preliminary estimation of parameters (among others conditional linearity approach) and iterative procedure of parameters estimation (the Newton-Rhapson algorithm). The results are compared to the well-known method of bankruptcy prediction, the Altman model based on stepwise discriminant analysis (Z-score).

Keywords: agribusiness, bankruptcy prediction, nonlinear models, production functions.

JEL classification: C50
AMS classification: 91G70

1 Introduction

Tools and techniques of prediction of corporate financial distress/failure bring together outputs of financial analysis related to the past economic performance of the business entity and the effort to estimate its future economic performance. The mentioned approach is being used within the corporate level, above all by banks and other private providers of external sources of capital and as well by government authorities in connection with provision of public subsidies. Especially the agricultural sector is the biggest recipient of public subsidies from EU budget via the Common Agricultural Policy and the focus of this policy is to provide sustainable development of agricultural enterprises across European Union.

Nevertheless, especially investment subsidies are intended to be provided only to financially healthy enterprises with further perspectives of their sustainable economic performance. So, the actual problem to solve is how to distinct the well performing enterprises from those likely to face serious financial problems. During several last decades numbers of different approaches were developed for classification of business failures of business entities; however, the emphasis was given mainly to industrial and financial enterprises.

The well know approaches in this area are mainly based on the stepwise discriminant analysis such as Z-Score model developed by Prof. Altman in 1960s for the sample of US companies. The similar approach was utilised for the development of IN bankruptcy – creditworthiness models specifically for the Czech business environment in late 1990s and early 2000s. Nowadays, there is a tendency to test new approaches in this area, for instance nonparametric methods such as the Data Envelopment Analysis or the artificial neural networks as non-linear statistical decision making modeling approach.

[1]Mendel University in Brno/Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, e-mail: david.hampel.uso@mendelu.cz
[2]Mendel University in Brno/Department of Business Economics, Zemědělská 1, 613 00 Brno, Czech Republic, e-mail: jan.vavrina@mendelu.cz
[3]Mendel University in Brno/Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, e-mail: jitka.janova@mendelu.cz

The involvement of production function (PF) within this mentioned area is a completely new approach considering the current state of art. The agribusiness case within this topic was chosen also due to fact that this industry produces very homogenous outputs/products. This assumption is considered by authors to be helpful within employing the PF based approach for the business failure classification.

1.1 Production functions

Typically, a PF relates the rate of output Q to the amount of two or more factors of production. In many cases PFs are used in a macroeconomic context (see for example [5]), nevertheless, it is a microeconomic concept. Very often a three-variable PF relates output to combinations of capital K and labor powers L is used, in the general form we can write

$$Q = f(K, L).$$

One of the most popular PFs used in empirical work is the Cobb-Douglas PF (see [4], [7]), which general form is

$$Q = \gamma K^{\alpha} L^{\beta}, \tag{1}$$

where γ, α and β are positive constants, with $\alpha < 0$ and $\beta < 0$. Generally, see [3], parameter γ relates to a level of technology, parameters α and β are output elasticities for labor and capital. By the sum $\alpha + \beta$ we can exhibit returns to scale. We can distinguish among three situations: for $\alpha + \beta > 1$ we have increasing returns to scale, for $\alpha + \beta < 1$ decreasing returns to scale and for $\alpha + \beta = 1$ constant returns to scale. For example in [3], there is stated modification of function (1) for $\beta = 1 - \alpha$ as

$$Q = \gamma K^{\alpha} L^{(1-\alpha)}. \tag{2}$$

This PF means constant returns to scale automatically. There are many generalizations of PF (1) and another PF. In [6], there is developed PF with constant elasticity of substitution (CES) of the form

$$Q = \gamma[\delta K^{-\rho} + (1 - \delta)L^{-\rho}]^{(-v/\rho)}, \tag{3}$$

where $\gamma > 0$, $0 < \delta < 1$, $v > 0$, $\rho \geq -1$. Note that γ is known as parameter of efficiency, δ as parameter of distribution, v as parameter of returns to scale and ρ as parameter of substitution.

2 Material and Methods

2.1 Data

The contribution is based on the secondary research and it is utilising the Database Amadeus of Bureau van Dijk as the ultimate source of economic data of corporates financial statements and the tool for the primary identification of economically distressed agricultural companies. The companies within the business failure sample (24 entities) are those who went bankrupt in year period 2008–2009. The searching strategy in Amadeus database to identify such companies was based on the classification of economic activities CZ NACE revision 2, namely section 01 – Crop and Animal Production. Another element of searching strategy was the activity status, when it was focused on enterprises with the business activity status bankruptcy, in liquidation or inactive (no precision). This mentioned activity status was revised for the respective business entity via officially issued documents in the Czech Business Registers.

The main problem related to the selection of the sample within the business failure evidence population was to identify agriculture companies, which went bankrupt after a period of no active agriculture production (e.g. entities which were involved in the process of transformation of agriculture cooperatives). The mentioned year period was chosen to cope with as much as possible present data in accordance with the current state of art concerning the EU Common Agriculture Policy and accessibility of complementary documents related to respective business failures via digitalized database of the Czech Business Register. On the other hand, the sample of well performing companies was set up by utilisation of the rating system MORE provided by the Database Amadeus including the agricultural companies with the highest awarded rating AAA/AA (21 entities).

The respective outputs and factors of production related to PFs were based on the data of financial statements prior to bankrupt of the respective company within the sample. The output of the PF was set up as the added value indicator. This indicator is for purposes of this contribution enumerated as

addition of interrelated indicators, namely profit/lost for current period, depreciation, taxation, interests paid and cost of employees. There is defined the factor of production capital as the addition of total shareholder funds and total liabilities and finally the factor of production labor as the number of annual working units within the respective year that represents annual full time job equivalent headcount.

2.2 Parameters estimation in nonlinear models

All introduced PF are nonlinear in parameters. Although it is possible to approximate a parameters estimation by linearization for some of PFs, we will use appropriate nonlinear techniques, described for example in [2]. We can express nonlinear model generally as

$$Y_i = f(\boldsymbol{x}_i, \boldsymbol{\theta}) + e_i, \quad e_i \sim \text{iid}(0, \sigma^2), \quad i = 1, \ldots, n,$$

where $\boldsymbol{x}_i$ is the vector $(1 \times k)$ of independent variables related to the value y_i, $\boldsymbol{\theta}$ is vector $(1 \times p)$ of parameters. The estimate of parameters we obtain (analogically to linear models) by minimizing

$$SSE(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \left[y_i - f(\boldsymbol{x}_i, \hat{\boldsymbol{\theta}}) \right]^2.$$

A minimization comes directly from the $SSE(\hat{\boldsymbol{\theta}})$ using a suitable numerical method of minimization. This approach is called the nonlinear least squares method (NLS).

Generally, an algorithm for parameters estimation is based on iterations and covers procedure for searching preliminary ("starting") estimates; a rule for "improving" iterations and a rule for the calculation ending. For setting preliminary estimates specialized procedures are developed. Often there is successful the choice based on depicting values, "sorting on the grid" or "random shooting". Improperly selected initial parameters can cause a significant prolongation or complete blocking of the calculation. The ending criterion is based on the iterations convergency (not on the number of iterations for example).

One of the suitable approaches may be to end the calculation, when the difference of the new SSE from the previous iteration varies less than a preselected small positive constant. Other criteria may be based on differences in parameter values or subsequent iterations resp. on the sum of their squares.

The most commonly used minimizing method is Levenberg-Marquardt algorithm (LMA) or its special cases – Gauss-Newton algorithm (GNA) or Newton-Rhapson algorithms (NRA). LMA gives parameters for which some function reaches its minimum. LMA is combination of the GNA and NRA. LMA is more robust than GNA, because convergence is achieved although preliminary parameters are far away from minimizing parameters. Unfortunately, LMA (as well as GNA and NRA) finds only local, not global, minimum.

2.3 Used datasets and methods

In the further text, we denote data for bankrupted companies as B2007 resp. B2008 for the year 2007 resp. 2008; for highly rated companies as A2007 resp. A2008 for the year 2007 resp. 2008. For the nonlinear estimation we use LMA, mostly with uniform choice of preliminary parameters. For testing differences of the PF parameters for bankrupted versus highly rated companies we use approach with dummy variable (see [2]). We denote this variable as D, $D_i = 0$ for a highly rated company and $D_i = 1$ for a bankrupted company. By this manner we obtain model for all companies in the particular year based on Cobb-Douglas PF (1)

$$Q = (\gamma_0 + \gamma_1 D) K^{(\alpha_0 + \alpha_1 D)} L^{(\beta_0 + \beta_1 D)} \tag{4}$$

and based on modified Cobb-Douglas PF (2)

$$Q = (\gamma_0 + \gamma_1 D) K^{(\alpha_0 + \alpha_1 D)} L^{[1 - (\alpha_0 + \alpha_1 D)]}, \tag{5}$$

where subscript 0 means parameters for highly rated companies and subscript 1 means parameters interpretable as increase or decrease of particular parameters for bankrupted companies. Calculations were made using R-based open source software Gretl 1.9.7 and computational system Matlab R2012a. Note that in Tab. 1–5 the symbol * means that the particular parameter is statistically significant in the model.

3 Results

We would like to model our data using CES PF (3) because of larger number of parameters and thus higher likelihood to distinguish between a group of bankrupted companies and companies in good condition then for other PFs. We assume agricultural sector labor-intensive, so CES should be adequate approach for it. The uniform choice of preliminary estimates was not successful, so the conditional linearity approach (see [2]) was used. Nevertheless, the estimation turned out well only for B2008 data. As we can see in Tab. 1, the estimate of the parameter δ is equal to 1, what means that CES PF is reduced substantially and using parameters v and ρ together has no sense.

γ	δ	v	ρ	R^2_{adj}
0.00000817	1.000*	1.546*	0.976	0.43

Table 1: Estimates of CES PF (3) parameters for B2008 data

According to this situation we use, for analysis of our task, PFs (1) and (2). Estimation of parameters was successful for all datasets. In Tab. 2 left chart we can see, that estimated parameters α and β are statistically significant and stable for highly rated companies. For bankrupted companies we obtain models with lower R^2_{adj} than for companies in good condition in general. In 2007 parameter α was significant but with strange estimate (greater than 1), parameter β was not significant. In 2008 both parameters weren't significant and differ from stable parameters of A2007 and A2008 datasets visibly.

It seems that parameter β in (1) is redundant for some datasets, so we try to compare them omitting this parameter. In Tab. 2 right chart, there are results for PF (2). Estimates of the parameter α are relatively close for all datasets except B2008. In this case, the estimate of the parameter α differ visibly and has no significance in the model. For all models and datasets, we perceive the parameter γ as necessary to be in a model, but useless for distinguishing among models.

Par.	B2007	B2008	A2007	A2008
γ	0.026	0.321	472.842	237.022
α	1.036*	0.873	0.539*	0.583*
β	0.312	0.430	0.321*	0.301*
R^2_{adj}	0.66	0.48	0.86	0.90

Par.	B2007	B2008	A2007	A2008
γ	34.419	324.144	42.309	31.877
α	0.630*	0.528	0.656*	0.680*
R^2_{adj}	0.66	0.49	0.84	0.89

Table 2: Parameter estimates for particular datasets. Left chart: PF (1), right chart: PF (2)

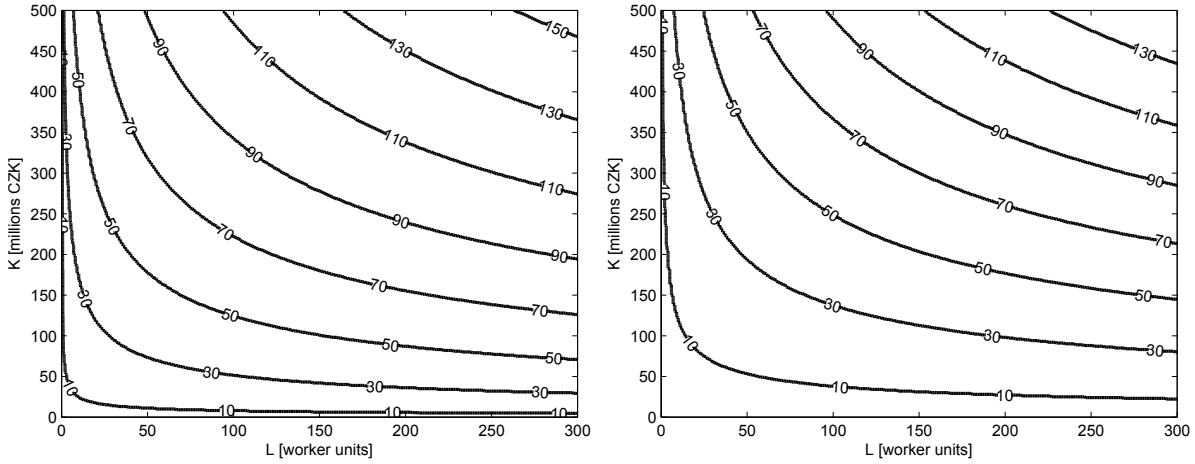


Figure 1: Contour plot of PF (1) for highly rated companies (left graph) and bankrupted companies (right graph) in the year 2008; depicted values of Q are in millions CZK

For illustration of estimated PFs we plot contours of PFs based on datasets A2008 and B2008 (see Fig. 1). In general, there is visible that highly rated companies use both production factors more effectively than bankrupted ones, what reflects in greater output for the same combinations of production factors.

Further, in the case of companies in good condition isoquants are denser for lower combinations of K and L than for higher combinations. In the case of bankrupted companies the reverse situation occurs.

Now we try to determine possible differences between parameters for highly rated companies and bankrupted companies in an exact way. In Tab. 3, there are parameters estimates for model (4). We deal with four submodels M1–M4 differing in absence of particular parameters (marked by the symbol "×") combined with two datasets – all companies in 2007 and in 2008. In the year 2007, there are no significant 1-indexed parameters. In following year, for submodels M1 and M2 significance cannot be approved, but for submodels M3 resp. M4 we have statistically significant parameters β_1 resp. α_1.

Par.	M1-2007	M2-2007	M3-2007	M4-2007	M1-2008	M2-2008	M3-2008	M4-2008
γ_0	472.841	472.841*	472.841*	472.841*	236.700	234.360	234.360	234.360
γ_1	-472.773	×	×	×	-236.499	×	×	×
α_0	0.539*	0.539*	0.539*	0.539*	0.584*	0.584*	0.584*	0.584*
α_1	0.445	0.445	×	0.445	0.312	-0.115	×	-0.115*
β_0	0.321*	0.321*	0.321*	0.321*	0.301*	0.301*	0.301*	0.301*
β_1	-0.0086	-0.0086	-0.0086	×	0.144	0.338	0.338*	×
R^2_{adj}	0.91	0.91	0.92	0.92	0.94	0.94	0.94	0.94

Table 3: Parameter estimates for model (4) using particular datasets and submodels

We can interpret this result in the sense, that in 2007 there weren't significant differences between companies being in good condition or bankrupt in 2008–2009 according to PF approach. But in 2008 significant differences appears, what means, that explored groups can be discriminated using PF (1) parameters. We assume, that the nonsignificance for submodels M1 and M2 is mainly caused by low number of companies.

Analogous technique is applied using model (5), see results in Tab. 4. Although parameter estimation results for model (2) demonstrate another behavior for the dataset B2008 than for others, there are no significance of α_1 in any case. It can be interpreted that the model (2) is insufficient to describe differences between bankrupted companies and companies in good condition.

Par.	M5-2007	M6-2007	M7-2007	M5-2008	M6-2008	M7-2008
γ_0	42.309	42.309	42.309	31.877	32.388	32.548
γ_1	-13.441	×	-13.441	528.141	×	-17.351
α_0	0.656*	0.656*	0.656*	0.680*	0.679*	0.679*
α_1	-0.014	-0.014	×	-0.246	-0.052	×
R^2_{adj}	0.90	0.90	0.90	0.93	0.93	0.93

Table 4: Parameter estimates for model (5) using particular datasets and submodels

Finally, we try to predict status of the company using PF approach. This prediction will be confronted with prediction obtained from Altman model (see [1]). The set of 5 bankrupted companies and 5 highly rated companies was chosen from datasets A2008 and B2008 in a random way. Resting data we denote as P-A2008 and P-B2008 and use them for estimation of model (1) parameters in both cases, see Tab. 5.

dataset	γ	α	β	R^2_{adj}
P-A2008	240.998	0.583*	0.301*	0.91
P-B2008	11.746	0.682	0.447	0.64

Table 5: Parameter estimates for model (1) applied on reduced datasets

It is visible, that for highly rated companies parameters remained almost the same like for non-reduced dataset A2008, but for bankrupted companies we can see differences. For 10 chosen companies we calculate predictions of product Q by both estimated PF (denote Q_0 for prediction using P-A2008 dataset and Q_1 for prediction using P-B2008 dataset) and classify them by comparison with real Q: If

$|Q - Q_0| < |Q - Q_1|$ we assume the company to be in good condition, if the opposite is true the company should tend to bankruptcy. The comparison of predictions success is outlined in Tab. 6 together with Altman model results coming from our previous work. We can see that the prediction was correct for 80 % of companies and it was at the same level of correct prediction as within the Altman model approach. Note that the Altman model missclasification was not so strict, the model denoted this two companies as "ambivalent financial situation", not coming to bankrupt.

Reality	highly rated companies		bankrupted companies	
Estimation	correct	wrong	correct	wrong
PF approach	3	2	5	0
Altman model	3	2	5	0

Table 6: Success of classification for randomly chosen companies provided by the PF approach and by the Altman model

4 Conclusions

PF approach seems to be adequate to distinguish between bankrupted and highly rated companies including prediction of this status. From the comparison of estimated PF contours in Fig. 1 we can conclude that highly rated companies use limited resources better than bankrupted companies. According to small-scale data, results can be considered as the first step for further analyses.

The Czech Agricultural producers do the economic activities within the framework of the current EU Common Agriculture Policy. This policy provides them financial subsidies which are mostly decoupled of the volume of production. So the economic performance of the respective business entities depends on the management of the factors of production and other scarce resources as well. The evidence of the Economic Agricultural Accounts held by the Czech Statistical Office provides the evidence of the net value added creation within the whole Czech agricultural industry. Subsequently the indicator net income represents the entrepreneurial revenue that consists both of incomes related to subsidies and income related to the current business activities. The subsidy system of the EU so cannot provide the ultimate financial sources for sustainable economic performance and that is why it is fully relevant to cope with the economic models to estimate and prevent the threat of economic distress situation.

Acknowledgements

This research is supported by the grant MSM6215648904.

References

[1] Altman, E. I.: Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, *Journal of Finance* **23** (1968), 589–609.

[2] Bates, D. M. and Watts, D. G.: *Nonlinear Regression Analysis and its Applications*. Wiley, Hoboken, New Jersey, 2007.

[3] Brennan, M. J. and Carroll, T. M.: *Preface to Quantitative Economics & Econometrics*. 4th Ed. South-Western Publishing, Cincinnati, 1987.

[4] Cobb, C. W. and Douglas, P. H.: A Theory of Production, *American Economic Review* **18** (1928), 139–165.

[5] Černý, J.: Comparison of Cobb-Douglas Production Functions of the Chosen Countries by Using Econometric Model. In: *Mathematical Methods in Economics 2011 – part I*, Professional Publishing, Praha, 2011, 107–112.

[6] Kmenta, J.: On Estimation of the CES Production Function, *International Economic Review*, **8** (1967), 180–189.

[7] Parkin, M.: *Economics*. Addison-Wesley Publishing Company, Reading, 1990.

Factors influencing the long-term unemployment level and development in the European Union

Jana Hančlová[1], Milan Šimek[2], Jiří Horák[3],

Abstract. The paper deals with the factors influencing the long-term unemployment rate in the European Union countries in 2001-2010. The applied factors were selected in compliance with the economics of labour market theory and cover three areas: economic environment (GDP growth in market prices, inflation), tax area (tax wedge), social benefit (net replacement ratio) and labour market flexibility (fixed-term contracts, part-time contracts). The panel model estimation has been implemented through a panel regression analysis for two homogenous groups (EU-15 countries and „new countries" – EU-12). The period preceding an economic crisis and a period of economic crisis have been investigated.
The results show that the labour market flexibility influences the long-term unemployment rate in the strongest negative way, then there is evident a positive impact of the tax wedge followed by the macroeconomic environment impact. The negative impact of economic growth operates in these groups with the same intensity, the tax burden in the EU15 countries is higher and vice versa with an increase in the share of part-time workers long-term unemployment decreases markedly in the EU12. The same positive economic growth will contribute to a deeper decrease of long-term unemployment. In the crisis period the positive impact of tax burden is reduced and we also monitor the increasing impact of the proportion of part-time jobs.

Keywords: long-term unemployment, European Union countries, labour market flexibility, panel regression, crisis.

JEL Classification: C23, E24
AMS Classification: 91B84

1 Introduction

The fundamental goal of the paper is to explore factors affecting the development of long-term unemployment. Four groups of factors are considered - macroeconomic environment (economic growth, inflation), the tax burden on labour (tax wedge), the social benefit (net replacement ratio) and the labour market flexibility (temporary contracts and part-time workers). Monitoring the impact of these factors will be further divided for the EU15 and EU12 because of homogeneity. Another key question of this paper is the impact of crisis on the development of long-term unemployment

The article is structured into five parts. Based on an introduction the theoretical concepts and long-term unemployment model are set in the second part. The third part specifies panel model of long-term unemployment by including dummy variables in the additive and multiplicative form. The fourth part deals with data analysis, the estimate of panel model and its verification and economic interpretation. The final part summarises the empirical results.

2 Model theory and concept

Empirical studies that are focused on examining the macroeconomic functioning of labour markets often try to identify factors affecting the level and changes in the level of unemployment. These analyses are usually focused on determining the natural rate of unemployment and an explanation of unemployment development after the second World War II in the OECD countries (Layard, Nickell, Jackman [9]).

In some studies the influence of long-term unemployment on other macroeconomic aggregates is analysed such as Lkaudes [10] examines the impact of long-term unemployment on inflation dynamics on the example of

[1] VSB-Technical University of Ostrava/Faculty of Economics, Department of Systems Engineering, Sokolská 33, Ostrava 1,701 21, Czech Republic, jana.hanclova@vsb.cz.
[2] VSB-Technical University of Ostrava/Faculty of Economics, Department of National Economy, Sokolská 33, Ostrava 1, 70121, Czech Republic, milan.simek@vsb.cz.
[3] VSB-Technical University of Ostrava/Faculty of Mining and Geology, Institute of Geoinformatics, 17. listopadu 15, Ostrava - Poruba, 70833, Czech Republic, jiri.horak@vsb.cz.

19 OECD countries, Tvrdoň [15] explores the relationship between economic cycle and unemployment in the countries of the Visegrad group using Beveridge curve , Katz [8] examines the long-term unemployment in the U.S. in 2007-2009 and Chapman and Kapuczinski [3] deal with the impact of recession on unemployment duration. In the model we use the assumption that there are factors that affect the level and change of unemployment rate. The investigated factors can be simply divided into **four groups** - the macroeconomic environment, the tax burden on labour, social benefits and labour market flexibility. The paper Fisher and Vltavska [4] also investigates effects of total factor productivity on economic output growth.

2.1 Macroeconomic environment (economic cycle, GDP growth, inflation of the subsection (automatically numbered)

Effect of GDP growth in market prices on long-term unemployment rate. Assuming an influence of Okun 's law it can be assumed that the growth of aggregate demand in the phase of economic expansion creates new jobs and thus a decline in unemployment, including long-term unemployment is expected. Tvrdoň [16] on the basis of a study of Abraham and Shimer shows that the average duration of unemployment has decreased for the majority of past business cycles but not as much as for unemployment rate, which has been explained by the dynamics of the labour market. Pisarides [12] describes a possible mechanism of unemployment extension to the overall functioning of the economy.

The relation of inflation and the long-term unemployment. Based on theoretical data based on the interpretation of Philips' curve the aim is to demonstrate trade off relationship between the rate of long-term unemployment and inflation in the monitored period. The relationship between general unemployment and inflation in the EU was examined by G. Popovich and Popovich J [13]. This paper analyses dynamics and interdependence of the ratio between inflation and unemployment rates in the EU for several referential periods (1998-2007, 2000-2009 and 2000-2006). Unemployment and inflation have a significant if inverse relation, both of which indicate the existence of Phillips regularity. The comparison of results has also confirmed the side hypothesis, that is, the regularity fades in case of economic shocks or periods of crises. The paper sets to prove that the long-term dynamics of these values in the EU is related to the findings of Phillips.

2.2 Tax burden on labour area

In this area the indicator called *tax wedge* has been used. The tax wedge on labour costs is defined as income tax on gross wage earnings plus employee and employer social security contributions, expressed as a percentage of total labour costs. This indicator is compiled for single people without children earning 67 % of the average earnings of a worker in business economy. The model assumes that the higher tax burden on labour will increase the rate of long-term unemployment. Tvrdoň [16] on the basis of research results indicates that the tax burden in this way affects the long-term unemployment. Compared to the impact on the unemployment rate, however, these effects were lower.

2.3 Social benefit area

The indicator called *net replacement rate* has been used in this area which is calculated as a share of net income share (after taxation) of unemployed person compared to the previous income after taxation. The closer to one a share is, the less income of unemployed people differ from those who are working. From this viewpoint, the higher value of this indicator can increase the long-term unemployment rate because of reduced incentives to work of unemployed. This situation is referred to as unemployment trap.

2.4 Labour market flexibility area

In this area the impact of *a share of part-time jobs and fixed contracts in the economy* on the long-term unemployment rate is examined. Both of the above forms through their implementation increase labour market flexibility in terms of removing barriers to the transition between jobs and the transition from economic inactivity or unemployment to employment (see Hančlová [5]). In both cases, greater use of these forms may reduce the unemployment rate, and also the long-term unemployment (see Pánková, [11]).

3 Data and methodological and specification issues

Based on theoretical concept of the model we proceed to the econometric formulation of the stochastic model. First of all we introduce the indications of following variables for cross-sectional units (countries EU27) $i = 1,2,\ldots,N$ monitored in period (year) $t = 1,2,\ldots,T$:

- LTU_{it} long-term unemployment the ratio of long-term unemployed more than 12 months on the number of economically active population in % for *i*-th country in year *t*;
- GDP_{it} percentage changes of GDP on previous period at market prices for *i*-th country in year *t*;

- INF_{it} annual change of Harmonized Index of Consumer Prices for *i*-th country in year *t*;
- TX_{it} tax wedge on labour costs as tax rate on low wage earners in % for *i*-th country in year *t*;
- NRR_{it} net replacement ratio for *i*-th country in year *t*;
- TCW_{it} percentage of employees with temporary contracts for *i*-th country in year *t*;
- PTW_{it} part-time workers in % of total employment for *i*-th country in year *t*.

We provide unemployment benefit replacement rates for a single worker. The calculations assume a worker, aged 40, who earns the average production worker wage.The selected file of indicators included annual data for 2001-2010. For investigation of crisis period impact a dummy variable $DCRIS_{it}$ ($DCRIS_{it} = 1$ for years 2008-2010, $DCRIS_{it} = 0$ otherwise) was introduced and two groups of countries were further investigated through variable DEU_{it}, which was specified as: $DEU_{it} = 1$ for new countries which joined the EU, $DEU_{it} = 0$ otherwise). Table 1 provides the indication of cross-sectional units and their division into groups.

group	country
$DEU_{it} = 0$ **countries EU15**	Belgium (BE), Denmark (DK), Germany (GE), Ireland (IE), Greece (EL), Spain (ES), France (FR), Italy (IT), Luxembourg (LU), Netherlands (NL), Austria (AT), Portugal (PT), Finland (FI), Sweden (SE), United Kingdom (UK)
$DEU_{it} = 1$ **counries EU12**	Bulgaria (BG), Czech Republic (CZ), Estonia (EE), Cyprus (CY), Latvia LV), Lithuania (LT), Hungary (HU), Malta (MT), Poland (PL), Romania (RO), Slovenia (SI), Slovak (SK)

Table 1 Specification of cross-section units and their divison into groups

Formulation of stochastic econometric linear model presents following equations:

$$LTU_{it} = \alpha_1 DCRIS_{it} + \alpha_2 DEU_{it} + \sum_{k=1}^{K} \beta_k X_{kit} + \sum_{k=1}^{K} \gamma_k (X_{kit} \cdot DCRIS_{it}) + \sum_{k=1}^{K} \lambda_k (X_{kit} \cdot DEU_{it}) + u_{it}, \quad (1)$$

where X_{kit} are individual explanatory variables $k = 1,2,\ldots,K$ and u_{it} are random errors which should be uncorrelated with the observed explanatory variables and with each other and have zero means.

Proposed model was estimated through Panel Generalized Least Squares (PGLS) method in software EViews. Version 7.1 [1]. This methods is also described in Wooldridge [17]. We provide estimation using panel GLS with cross-section weights for removing cross-section heteroscedasticity, which is investigated in Hančlová [6]. We can also generally expect the errors $(u_{it} : t = 1,2,...,T)$ serially correlated and therefore we used PGLS estimation method. We also tested fixed or random cross-section, resp. time effects but they were not significant. These effects describes Ivaničová, et al. [7]. Cerný [2] deals with linear regression with special data.

4 Empirical results

First, we analyse of input time series in accordance with paper Rublíková [13]. We introduce dummy variables *DCRIS* in additive and multiplicative form for explanation the change in times of crisis. We compare the cross-section parameters differences between the EU15 and EU12 groups through a dummy variable *DEU* again in additive and multiplicative form. Figure 1 shows the level and development of long-term unemployment. The picture and other descriptive statistics showed that in the EU15 group of countries the highest median long-term unemployment was in Germany, Spain and Portugal, where also rose rapidly. In the EU12 the average long-term unemployment was the highest in the Slovakia, Poland and in Bulgaria. Results of analysis of other indicators are not presented because of the scope of the paper.

When estimating panel model for 27 countries in 2001-2010 explanatory variables of inflation (INF_{it}) and net replacement ratio (NRR_{it}) were found statistically insignificant, thus excluded from the model. In addition, to explain the model it was more appropriate to use lagged explanatory variables GDP_{it-1} and TCW_{it-1}. The final estimation of the model (1) by PGLS method is set in the table 2. Determination coefficient R-squared was 0.68. We provide the multivariate extensions of the Jarque-Bera residual normality test which compares the third and fourth moments of the residuals to those form the normal distribution. JB statistics was 4.29 with significance 0.116, ie. we do not reject the null hypothesis of normal distribution of residual components at the 5% significance level.

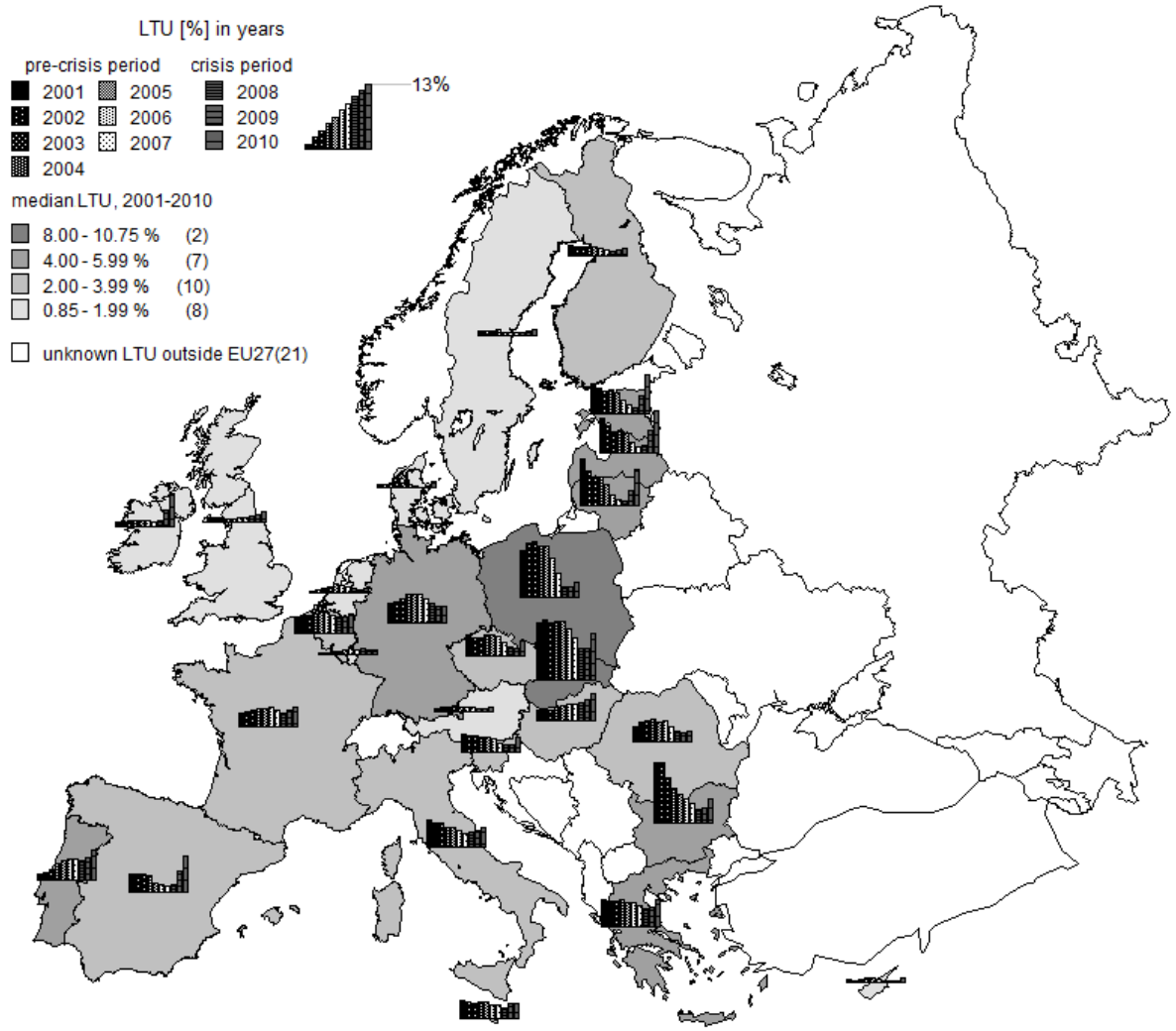


Figure 1 Long-term unemployment development in countries EU15 and EU12

Dependent Variable: LTU, *Method*: Panel EGLS (Cross-section weights), *Sample* (adjusted): 2002 2010
Periods included: 9, *Cross-sections* included: 27, *Total panel* (unbalanced) observations: 238

Variable	*DCRIS*	*DEU*	*GDP(-1)*	*TX*	*TCW(-1)*	*PTW*	*DCRIS_GDP(-1)*	*DCRIS_TX*	*DCRIS_TCW(-1)*	*DEU_TX*	*DEU_TCW(-1)*	*DEU_PTW*
Coefficient	2.397	4.002	-0.105	0.099	0.035	-0.066	-0.085	-0.086	0.048	-0.036	-0.097	-0.129
Std. Error	0.25	0.504	0.031	0.004	0.009	0.006	0.038	0.011	0.02	0.006	0.016	0.05
t-Statistic	9.592	7.937	-3.437	26.56	3.940	-10.55	-2.259	-7.716	2.367	-6.155	-5.97	-2.609
Prob.	0.0000	0.0000	0.0007	0.0000	0.0001	0.0000	0.0248	0.0000	0.0188	0.0000	0.0000	0.0097

Table 2 Final estimationof long-term unemployment model
Source: EViews and author calculations

The estimations of average long-term unemployment (hereafter *LTU*) in the model can be divided into *four groups* in order of better interpretation according to the dummy variables *DCRIS* and *DEU.*

Group of countries EU15 and pre-crisis period 2001-2007 ($DCRIS_{it} = 0 \wedge DEU_{it} = 0$)

$$\hat{LTU}_{it} = -0.105GDP_{it-1} + 0.099TX_{it} + 0.035TCW_{it-1} - 0.066\,PTW_{it} \tag{2}$$

GDP growth by 1% in previous year enables a decrease of *LTU* by 0.105 % ceteris paribus, tax wedge growth by 1 % contributes to the increase of *LTU* by 0.099 %, one percent growth of employees with temporary contracts in previous period causes an increase of *LTU* by 0.035% a higher share of part-time workers on total employment by 1 % results in a decline of *LTU* by 0.066 %.

Group of countries EU12 and pre-crisis period 2001-2007 ($DCRIS_{it} = 0 \wedge DEU_{it} = 1$)

$$\hat{LTU}_{it} = 4.001 - 0.105GDP_{it-1} + 0.063TX_{it} - 0.062TCW_{it-1} - 0.196\,PTW_{it} \quad (3)$$

One percent GDP growth contributes to the same decrease of *LTU* as in EU 15 countries in 2001-2007, tax wedge growth by 1 % results in an increase of *LTU* by 0.063 %, one percent growth of employees with temporary contracts in previous period influences *LTU* development negatively (-0.06 %) and an increase of share of part-time workers on total employment by 1 % causes a decrease of *LTU* by 0.196 %, which is greater decrease compared to the EU15.

Group of countries EU15 and crisis period 2008-2010 ($DCRIS_{it} = 1 \wedge DEU_{it} = 0$)

$$\hat{LTU}_{it} = 2.397 - 0.190GDP_{it-1} + 0.013TX_{it} + 0.083TCW_{it-1} - 0.066\,PTW_{it} \quad (4)$$

GDP growth by 1 % enables to decrease an average *LTU* by 0.190 % in the following year which is more radical decrease in comparison with the pre-crisis period. Tax wedge growth by 1 % increases *LTU* by 0.013 % which is more in comparison with EU15 and EU12 in pre-crisis period. One percent growth of employees with temporary contracts results in an increase of long-term unemployment by 0.083 % in the following year and an increase of share of part-time workers on total employment by 1 % results in decline of long-term unemployment by 0.066 % which is for this group of countries the same as before crisis.

Group of countries EU12 and crisis period 2008-2010 ($DCRIS_{it} = 1 \wedge DEU_{it} = 1$)

$$\hat{LTU}_{it} = 6.399 - 0.190GDP_{it-1} - 0.017TX_{it} - 0.014TCW_{it-1} - 0.196\,PTW_{it} \quad (5)$$

Increase of GDP by 1 % contributes to the same reduction of *LTU* as for countries and one percent increase in tax wedge causes reduction of *LTU* by 0.017 % in line with expectations. Effect of growth of employees with temporary contracts by 1 % causes as in pre-crisis period the decline of *LTU* but with less intensity (-0.014 %). An increase of the share of part-time workers on total employment by 1 % means reduction of *LTU* by 0.196 % which is the same as for this group of countries in crisis-free period.

5 Conclusions

Comparison of the results of long-term unemployment estimation can be summarized and assessed in terms of those groups of factors, by examining the differences between the group of EU15 and EU12. Comparison of the results of long-term unemployment development can be added to the previous development in 2001-2007 for a period of crisis.

In accordance with the theoretical background the long-term unemployment was affected by:

- macroeconomic environment (due to negative economic growth, which was for the EU15 and the EU12 the same and more intensive in period of crisis);
- tax burden on labour (usually a positive effect which was higher for the EU15 group of countries and in the crisis period decreased in both groups);
- flexibility of labour (the negative impact on part-time workers in % of total employment which is stronger in the EU12, and not influenced by the crisis; effect of temporary contract factor is in the EU12 in line with expectations negative and during the crisis is less intensive but for a group of EU15 we observe a positive impact on long-term unemployment).
- Group of social benefits statistically did not influence significantly on long-term unemployment as well as the inflation.

In terms of group comparison of EU15 and EU12 it was found that the negative impact of economic growth operates in these groups with the same intensity, the tax burden in the EU15 countries is higher and vice versa with an increase in the share of part-time workers long-term unemployment decreases markedly in the EU12. An interesting finding was that the labour flexibility indicator (temporary contracts) acts on long-term unemployment in the EU15 positively and in EU12 negatively.

By comparing the 2001-2007 period and crisis period it was found that the same positive economic growth will contribute to a deeper decrease of long-term unemployment. In the crisis period the positive impact of tax burden is reduced and we also monitor the increasing impact of the proportion of part-time jobs.

The results obtained are consistent with economic theory and researched professional publications and has been verified for two specific groups of countries in the years 2001-2010 through panel model of long-term unemployment.

Acknowledgements

The paper was funded by Student Grant Competition of VŠB-TU Ostrava within the project SP2012/62.

References

[1] EViews. *Eviews 7 User'Guide II.* Quantitative Micro Software, Irvine, 2009.
[2] Černý, M., and Rada, M. On the Possibilistic Approach to Linear Regression with Rounded or Interval-Censored Data. *Measurement Science Review* **2** (2011), 34-40.
[3] Chapman, B., and Kapuscziski, C.: Avoiding Recessions and Australian Long-Term Unemployment. *Discussion Paper – The Australia Institute* **49** (2000).
[4] Fisher, J., and Vltavska, K. Total factor productivity measurement based on labour and capital services. *Politicka ekonomie* **5** (2011), 599-617.
[5] Hanclova, J. Panel Modelling of Globalization on Government Expenditures for the Selected New EU Countries. In: 1*0th International Conference of Liberec Economic Forum.* Technical University of Liberec, Liberec, 2011, 161-170.
[6] Hanclova, J., and Doucek, P. Education and ICT Sector in the EU (Panel-National Application). In: *Proceedings of the 8th International Conference Efficiency and Responsibility in Education 2011.* Czech University of Life Sciences Prague, Prague, 2011, 84-92.
[7] Ivaničová, Z.,and Chocholatá, M., and Surmanová, K. *Ekonometrické modelovanie*, Ekonóm, Bratislava, 2012.
[8] Katz, L. F.: Long-Term Unemployment in the Great Recession. *Testimony for the Joint Economic Committee U.S. Congress* **April** (2010), 1-13.
[9] Layard, R.. and Nickell S., and Jackman R.: *Unemployment. Macroeconomics Performance and the Labour Market.* Oxford University Press, New York, 2005.
[10] Llaudes, R.: The Phillips Curve and Long Term Unemployment. *Working Papers Series - European Central Bank* **441** (February 2007), 1-49. http://www.ecb.int/pub/pdf/scpwps/ecbwp441.pdf
[11] Pánková, V. Tobinovo Q – teorie a aplikace. *Politicka ekonomie* **5** (2005) , vol. LIII, no. 5, 601–608.
[12] Pissarides, Ch. A. Loss of Skills During Unemployment and the Persistence of Employment Shock. *The Quarterly Journal of Economics* **4** (1992), 1371-1391. http://qje.oxfordjournals.org/content/107/4/1371.abstract
[13] Popovic, G., and Popovic, J.: Inflation and unemployment in the EU: comparative analysis of Phillips regularity. In: *Proceedings and papers - Sarajevo Business and Economics Review.* Faculty of Economics, Sarajevo, 2010.
[14] Rublíková, E. Statistical Analysis of Chosen Macroeconomic Variables. In: *25th International Conference on Mathematical Methods in Economics.* VSB-Technical University of Ostrava, Ostrava, 2007, 310-315.
[15] Tvrdoň, M.: The business cycle and unemployment: empirical evidence from the Visegrad group countries. International. *Journal of mathematical model and methods in applied science **3*** (2001), 679-687.
[16] Tvrdoň, M.: *Institucionální rámec fungování trhu práce v kontextu ekonomické konvergence a přijetí společné měny. Aplikace na země Visegrádské skupiny.* Institut vzdělávání SOKRATES, Brno, 2011.
[17] Wooldridge, J. M. *Econometric analysis of cross section and panel data*. MIT Press, Cambridge, 2010.

Data sources

Net replacement ratio
http://www.neujobs.eu/sites/default/files/publication/2012/01/Unemployment%20Replacement%20Rates%20Dataset%20-%20Van%20Vliet%20&%20Caminada%20-%202012%20(1).pdf

European Commission. *Eurostat* [online]. 2012 [cit. 2011-10-05]. Statistics Database. Dostupné z WWW: <http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database>.

- *Long-term unemployment* - Annual average, by sex (%) [une_ltu_a]
  http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=une_ltu_a&lang=en
- *Tax wedge on labour costs* [earn_nt_taxwedge]
  http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=earn_nt_taxwedge&lang=en
- *Employment (main characteristics and rates)* - Annual averages [lfsi_emp_a]
  http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=lfsi_emp_a&lang=en
- *HICP (2005=100)* - Annual Data (average index and rate of change) [prc_hicp_aind]
  http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=prc_hicp_aind&lang=en
- *GDP and main components - volumes* [nama_gdp_k]
  http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=nama_gdp_k&lang=en

Do broker/analyst conflicts matter? Detecting evidence from internet trading platforms

Jan Hanousek[1], František Kopřiva[2]

Abstract. We analyze the potential conflict of interest between analysts and brokers associated with each other. In contrast to the existing literature we do not analyze prediction accuracy and/or biases in analyst recommendations. Instead we focus our analysis on brokers and examine whether their behavior systematically differs before and after investment recommendations are released. The evolution and dynamics of brokers' quotes and trades are used to test for systematic trading patterns around the release of one's own investment recommendation. In the model we control for brokers' responses to other investment advice and employ a SUR estimation framework. Data from the Prague Stock Exchange are used to demonstrate our methodology. Finding significant and systematic differences in brokers' behavior, we conclude that misuse of investment recommendations is widespread.

Keywords: dealers' market, emerging markets, informed trading, investment recommendations, trading systems.

JEL Classification: G14, G15, P34
AMS Classification: 62P20, 62J12, 62J05

1 Introduction

The integration of brokerage and analytical services on a stock market creates conditions for a particular type of conflict of interest. This conflict of interest can manifest itself in two ways: First, analysts may have an incentive to issue biased recommendations. Second, even if investment advice is unbiased, associated brokers may possess this information well before the other market participants and use it to their advantage. While the vast majority of theoretical and empirical research on investment recommendations focus on the first issue, i.e., analyst behavior, in this study we explore the second mechanism using publicly available high-frequency data.

We treat the recommendation of a particular analyst as new information that affects the decision-making of all market participants. Associated brokers, however, may have access to this information before it is released to the public and may thus possess an informational advantage over the rest of the market. The primary goal of our paper is to detect the misuse of this advantage by analyzing brokers' trading behavior prior to and after the investment recommendation is issued, with a particular emphasis on their responses to their own recommendations. If there is no conflict of interest, we should not see any **systematic** trading patterns a few trading days before or after an associated analyst issues a recommendation.

Contrary to the existing literature, we don't perform our analysis on restricted-access regulatory data, but use instead publicly available high-frequency data from trading platforms. This way, we do not only introduce a new approach to the analysis of investment recommendations but also overcome the problem of missing data, as the evolution of quotes and trades should very well replace the (often missing) regulatory information on stock inventories, portfolio structure, proprietary trading profits, etc.

1.1 Effect of Investment Recommendations and Emerging Markets

Interestingly, most of the papers analyzing the effects of investment recommendations have been conducted on data from developed capital markets such as that in the U.S., where regulation is quite strict and requires a separation of brokerage and investment banking activities. Therefore, it is not surprising that the results of these studies generally do not support the hypothesis that investors are systematically misled by investment recommendations. It would be a mistake, however, to extend these findings to emerging markets for several reasons, including:

1. Emerging capital markets are typically not subject to a high level of regulation;

[1] CERGE-EI, Charles University and the Academy of Sciences, Prague, CERGE-EI, P.O. Box 882, Politických veznu 7, Prague 1, 111 21, Czech Republic, jan.hanousek@cerge-ei.cz.
[2] CNB and CERGE-EI, Charles University and the Academy of Sciences, Prague, CERGE-EI, P.O. Box 882, Politických veznu 7, Prague 1, 111 21, Czech Republic, frantisek.kopriva@cerge-ei.cz.

2. Due to the smaller size of emerging markets, brokers have strong market power and substantial latitude for price manipulation;
3. Small investors are inexperienced and unaware of a possible conflict of interest.

In addition, Girard and Biswas [1] show that compared to developed markets, emerging markets exhibit greater sensitivity to the unusual volumes associated with the release of new information such as investment recommendations. Moreover, control over large market participants is usually limited as the regulatory authority often does not collect data on proprietary trading, broker inventories, etc. Overall, the problem of the misuse of investment recommendations and price manipulation could be more severe in emerging markets.

Unfortunately, the existing literature studying the effect of investment recommendations in the context of emerging markets is very limited. Moreover, the existing studies, likely due to the lack of regulatory data, analyze interactions between analyst recommendations and price change. For example, Moshirian, Ng, and Wu [4] in their sample of 13 emerging countries show that stock prices react strongly to stock analyst recommendations. They also report a stronger positive bias in analyst recommendations and revisions in emerging markets compared with that in developed markets. Similarly, Kiymaz [3] analyzed the effects of stock market rumors related to information release at the Istanbul Stock Exchange. He found that positive and significant abnormal returns are observed in the days prior to the publication date and negative yet insignificant returns are observed in the post-publication period. This supports our view of emerging markets, especially the possible existence of information leak and/or strategic trading around the time when recommendations are released.

In this study we use data from the Prague Stock Exchange in the Czech Republic to demonstrate our methodology. The Prague Stock Exchange represents a typical electronic dealers market, in which market makers play a dominant role in affecting the price for a short time interval as well as for a longer period (e.g., Hanousek and Kopriva [2]).

1.2 Methodology – Using Stock Quotes to Analyze Broker Behavior

Our aim is to capture and analyze brokers' trading behavior, particularly focusing on the dates around when associated analysts issue investment recommendations. Recently, several trading platforms enable participants to see real time quotes/positions of each active broker/market maker. So, there exist high-frequency trading data for each broker.

In general, brokers' trading positions are well described by their buy/sell statistics and by their positions on the bid and ask sides. In our analysis, we omit all cross trades, mainly because of their different nature (for example, some are used to set up standard operations such as the leveraged trading of a broker's client) and we consider only mandatory trades. For capturing and summarizing a broker's position on a bid or ask we use the relative distance from the best quotes and average the rank on the bid or ask. Since the original trading data are collected at a high frequency, for daily versions we need to compute the time-weighted averages of these variables.

From the methodological point of view, in our analysis we follow three key steps:
1. Construct statistics that summarize the daily behavior of a given broker from publicly available high-frequency data.
2. Use these statistics as dependent variables in seemingly unrelated regressions (SUR) with information about the timing and direction of investment recommendations used as a regressor.
3. Assess and comment on the differences in trading behavior of brokers before and after their own recommendations as well as the differences in reacting to one's own as opposed to external recommendations.

In the first step, we summarize brokers' trading behavior at a daily frequency in the following variables (computed for each share).

a) $Buy_{j,t}$ and $Sell_{j,t}$ = total number of mandatory buys and sells (in lots) by each broker j on trading day t.
b) Time-weighted percentage difference from the best bid (ask), computed for broker j on trading day t.
c) Time-weighted daily average rank on the bid (ask) of broker j.

The variables defined in a) through c) above do not only reflect the trading behavior of a particular broker but also allow for comparison with the behavior of other brokers. We are not only interested in the change of the behavior of an individual broker before and after a particular recommendation is issued, but also in whether his change of behavior is significantly different from the behavior of other brokers on the market.

Since we know the exact timing of each recommendation, we can link them with brokers' trading behavior and analyze the differences around the releases of the recommendations. The reaction to a particular recommendation differs depending on whether it is a positive (BUY) or negative (SELL) recommendation and also how the particular recommendation compares with his previous recommendation. Therefore, for each broker we define

several 0/1 indicators (dummy variables) defining a neighborhood of *k* trading days before and after the release of the recommendation by associated analysts. The choice of *k* trading days allows us to see whether there is a reaction and if so how long it lasts (as a baseline we use *k* = 5 and 10, but we also consider *k* = 15 and asymmetric windows for robustness checks).

To analyze effect of investment recommendations on the trading behavior of broker j let us consider the following specification:

$$trading_{jt} = \eta_j + \sum_{r=type\ of\ reccomendation} \Big[\beta_j \mathbf{B}^{r}_{j,own_t} + \alpha_j \mathbf{A}^{r}_{j,own_t} + \gamma_j \mathbf{B}^{r}_{j,other_t} + \delta_j \mathbf{A}^{r}_{j,other_t} + \varphi_j \mathbf{B}^{r}_{ext_t} + \phi_j \mathbf{A}^{r}_{ext_t}\Big] + \varepsilon_j \qquad j=1,...,J,\ t=1,...,D, \tag{1}$$

where the variable $trading$ stands for all trading proxies defined in a) through c), i.e., the number of buys and sells; the percentage difference from the best bid and ask; and the average rank of the bid and ask. As mentioned above, as baselines we will use time windows of five and ten trading days. Here, the dummy variables $\mathbf{B}^{r}_{j,own}$, $\mathbf{B}^{r}_{j,other}$, and $\mathbf{B}^{r}_{ext}$ are equal to one for ten trading days **before** the particular recommendation (own, other and external brokers) has been released. We keep the notation as it is above, where the subscript own is used when the associated analyst posted the recommendation, the subscript other is used when at least one of the other brokers posted the recommendation, and the subscript outside is used when an external analyst posted the recommendation. Similarly, $\mathbf{A}^{r}_{j,own}$, $\mathbf{A}^{r}_{j,other}$, and $\mathbf{A}^{r}_{ext}$ are equal to one for ten trading days **after** the particular recommendation was made public.

As specified in **(1)** we analyze the trading patterns by regressing each particular trading variable on a set of dummies representing the timing of all of the types of recommendation issued by all the kinds of analyst. There exists one potential problem, though. The analysis of trading data could indicate interactions between associated analysts and brokers, but might not detect which came first. If a broker's trading was primarily based on previous knowledge of an associated analyst's recommendation (more likely) or if the recommendation was released in order to maintain the brokers' inventories (less likely), there could be an endogeneity problem related to the timing of the recommendation.

For an estimation of the empirical specification we employ a seemingly unrelated regression (SUR) setup, where for a given share we estimate equations for all brokers *j =1,..,J* together. This approach provides more efficient estimates of the parameters of interest by using cross-equation correlations caused by, e.g., common exogenous shocks affecting all brokers such as a change in market trends, the common view of the particular stock, etc. Since we estimate specification (1) over the whole sample period, we control for interference between various recommendations (released at a similar time) and the heterogeneous shocks affecting the behavior of all brokers.

2 Data and Descriptive Statistics

For our analysis, we use information about all investment recommendations during the period 2003–2008, which are publicly available online at www.ipoint.cz, together with high frequency data about broker activity on the SPAD trading system of the Prague Stock Exchange (PSE), also publicly available online at www.akcie.cz. In the analysis we only use data on blue chip stocks, i.e., shares from the top-tier trading segment.

The high-frequency trading data consists of all SPAD trades and all SPAD quotes with an identification of brokers/market makers for all stocks traded during the time span 10 February 2004 to 31 December 2008.

Further, we divided the investment recommendations by the type of issuer:

1. investment firms that act also as brokers on the PSE (11 firms),
2. all other external investment analysts/firms who posted at least one investment recommendation during the analyzed time span.

3 Results

The main goal of this study is to find out whether brokers on the stock market misuse the potential informational advantage stemming from their association with analysts. This misuse could manifest in behavior different from other market participants. Even though studying individual patterns of significance and the direction of the coefficients for each stock and broker pair could be interesting from a regulatory point of view, we instead summarize these patterns in a comprehensive way to see the overall picture. During the studied period the Prague Stock Exchange was mostly on an upside trend, so we present here only the results for positive recommendations.

By analyzing the systematic patterns in estimated coefficients, we can answer questions regarding the existence of informational advantage and how it was used.

1. *Are the timing of recommendations unknown to associated brokers?*

This question can be addressed by looking at whether the estimated coefficients $\hat{\beta}_j$ (i.e., a systematic shift before the release of one's own recommendations) are significantly different from zero. For the 10-day window we see significant coefficients in about 28% of the cases for rank positions, 32% of the buy/sell measures, and about 42% of the quotes. For the 5-day window the results are even stronger. We observe significant coefficients in about 30% of the cases for rank positions, 36% of the buy/sell measures and about 44% of the quotes.

2. *Is broker trading behavior (around the time of the investment recommendation) consistent with the recommendation issued by associated brokers?*

Table 1 summarizes the consistent and inconsistent trading behavior of all brokers using a detailed combination of the possible outcomes before and after release of the recommendation.

	Time window	Buy and sell	Bid and ask difference	Bid and ask rank
All 15 companies	5 days	27%	22%	48%
	10 days	28%	27%	21%
Local large companies	5 days	34%	26%	19%
	10 days	30%	32%	29%
Cross-listed companies	5 days	23%	21%	13%
	10 days	33%	27%	15%

Table 1. Broker behavior systematically against his recommendation – sensitivity analysis, across various share groups and trading indicators (positive recommendations, 5- and 10-day windows)

The table contains a simple counting of each broker and stock of trading patterns, which are in line with or against the broker's own recommendation. The summary is based on the significance of the coefficients in specification (1). Below each behavioral proxy we present a ratio of all cases (broker and stock) in which we see the broker's behavior systematically contradicting his recommendations. The group of local large companies consists of O2 , CEZ, UNI, PM, Zentiva, and KB. Cross-listed companies are represented by CME, EB, ORCO and NWR.

From **Table 1** it is clear that the discussed inconsistency between investment recommendations and brokers' trading patterns is not just a coincidence. Very similar results are obtained for both time windows. As the inconsistent combinations indicate a recurring misuse of informational advantage stemming from affiliated analyst recommendations, **Table 1** demonstrates how the broker's behavior systematically contradicts his recommendations across various stock groups and trading proxies.

4 Conclusion

In this study we suggest an innovative approach to testing the potential conflict of interest between analysts and traders, specifically the potential misuse of investment recommendations. In contrast to the main stream of research associated with investment recommendations, we do not analyze the behavior of analysts; neither do we estimate their forecast error nor test if they behave strategically. Instead, we take their investment recommendations as given, including the timing, and analyze the behavior of associated brokers around the time of the release of a recommendation.

The difference in our approach also lies in the use of different data sources. We do not use data from regulating authorities as other studies do. We instead rely on high-frequency data from internet-based trading platforms that allow us to identify the intraday behavior of large brokers (market makers). We define time-weighted variables that summarize a broker's daily trading pattern, including his position on the bid and ask sides.

Our methodological approach is demonstrated on trading data from the Prague Stock Exchange. Results confirm that on this market the above-mentioned conflict of interest exists and is quite severe. Assuming that this

result can be generalized to all emerging stock markets, our findings support a need for the regulation of investment recommendations.

Acknowledgements

The authors would like to thank Štepán Jurajda, Evžen Kocenda and Peter Zemcík for helpful comments. Earlier versions of this paper were presented at CERGE-EI, UPENN, and various conferences; we acknowledge comments by presentation participants. The research for this study was supported by a GACR grant (403/11/0020). The views expressed are those of the authors and do not necessarily reflect the position of any of the affiliated institutions.

References

[1] Girard, E., and Biswas, R.: Trading Volume and Market Volatility: Developed vs. Emerging Stock Markets. *The Financial Review* **42** (2007), 429–452.

[2] Hanousek, J., and Kopriva, F.: Detecting Information-driven Trading in a Dealers' Market. *Czech Journal of Economics and Finance* **61** (2011), 204–229.

[3] Kiymaz, H.: The Stock Market Rumours and Stock Prices: A Test of Price Pressure and Size Effect in an Emerging Market. *Applied Financial Economics* **12** (2002), 469–474.

[4] Moshirian F., Ng D., and Wu E.: The Value of Stock Analysts' Recommendations: Evidence from Emerging Markets. *International Review of Financial Analysis* **18** (2009), 74–83.

Does financial support from the EU structural funds has an impact on the firms' performance: evidence from Estonia

Jelena Hartsenko[1], Ako Sauga[2]

Abstract. The purpose of this paper is to provide an empirical analysis to look for causal effects between government grant and performance indicators at the firm level. To analyze the impact of grants on the firms' performance we employ a difference-in-differences estimator. Such approach can be used separately for the firms that received the grant in different year. The second method we use is a panel-data framework. Results show different impact of various type of grant on firm's performance by economic activities.

Keywords: difference-in-difference, panel data, Estonia, government grant

JEL Classification: D24, H25, O32
AMS Classification: 62P20

1. Introduction

The main goal of government financial support is to promote business in order to accelerate economic growth. The financial support from government has been an important instrument of industrial policy especially in the transition countries. The Estonian Government has introduced different programs in order to support local enterprises. There are 15 programs available for SME for example the start-up and development grants, export grants, grants for research and development (R&D) activities and technology investments by companies.

In 2000 was established Enterprise Estonia (EAS) to promote business and regional developments in Estonia. Enterprise Estonia is one of the largest institutions within the national support system for entrepreneurship. Most of the EAS programs and grants combined with the co-financing from the EU structural funds. Since 2008, Enterprise Estonia has financed 960 projects with a total of 7.4 million euros from the European Social Fund [8]. A recent study by KPMG has revealed that between 2007 and 2009 Estonia was the most successful CEE country in utilizing grants from the European Union's structural and agricultural funds.

The stages of development of enterprise sector in Estonia a little differ from other transition economies. The liberalization and decentralization of the economy accompanied by lax monetary and fiscal policies contributed to overly fast growth of the number of firms. As a result of rapid privatization 90% of all Estonian enterprises were privately owned by 1995 and there are 58347 enterprises in Estonia in 2010. Within the group of small and medium sized enterprises (SMEs), a vast majority of the enterprises (88.9%) are micro enterprises, employing less than 9 persons. So, the typical Estonian firm is a micro firm. There are about 9.1% small and 1.8 % medium enterprises. On average, it was found that there is an increase of about 6% enterprises annually in the periods 1995-2010. At the moment there are approximately 44 SMEs per 1000 inhabitants and 81.91% of total employment is provided by SMEs.

Thus, it is important to evaluate the effectiveness of government financial assistance. Girma et al. [10] has argued that the evidence on effectiveness is mixed due the indicators for measuring effectiveness have been very various. Masso et al [15] pointed out that the effectiveness has been measured as the better use of technology, higher productivity, higher probability of firm survival or creating new jobs, whether subsidies crowd out or add to the firm's private expenditures (e.g. concerning R&D subsidies).

It is now widely acknowledged that increases in productivity are the main source of long-run economic growth. Lack of growth-oriented firms is argued to be one of the main obstacles to economic growth and increase in employment. Therefore the main hypothesis of the paper is that financial support from the government has positive impact on the firms' performance and accelerates economic growth through improved productivity. The present paper contributes to the literature on evaluating the effectiveness of government grants given to enterprises in Estonia. To analyze the impact of grants on the firms' performance we employ a difference-in-differences and a panel-data framework. We therefore evaluate the effectiveness of grants and

[1] Tallinn University of Technology, The Department of Economics and Business Administration, Akadeemia tee 3, jelena.hartsenko@tseba.ttu.ee
[2] Tallinn University of Technology, The Department of Economics and Business Administration, Akadeemia tee 3, ako.sauga@tseba.ttu.ee

distinguish the effects of different types of grants on firms' performance. Data of the firms, both supported and unsupported, are collected from Enterprise Estonia and Estonia Central Commerce Register from 2004 to 2010.

The remainder of the paper is organized as follows. Section 2 gives an overview of empirical research in the existing literature. The following section 3 contains data and methodology. In the section 4 are performed the econometric analysis and discussion. The final section includes conclusions.

2. Literature review

It is important to evaluate the effectiveness of the governmental assistance. There are many empirical studies estimating the impact of government grants to enterprises in different countries (Bergström [2], Almus [1], Crepon and Duguet [5], Girma et al [10], Ege [7], Sissoko [16], Criscuolo et al [6]). Estonian evidence on the determinants of firm growth is scant. The main studies in Estonia were done by Masso and Vildo [14] and Lukason and Masso [13]

In empirical literature the definition of efficiency has included many different domains and opportunities – the effectiveness has been defined through improved usage of technology, increased productivity or whether it has increased the probability of enterprise survival (Masso and Vildo [14]). Some of the empirical studies are briefly listed below. Bergström [2] showed in case of Sweden that subsidization is positively correlated with growth of value added and that productivity of the subsidized firms seems to increase the first year after the subsidies were granted. Almus [1] found from analysis of German data using parametric selection approach that firms receiving assistance perform better in terms of employment growth over a six year period. Crepon and Duguet [5] showed in case of French data with propensity score matching that start-up subsidies increased significantly the survival of the firms created by former unemployed people; and the allocation of subsidies acted as a screening process improving the performances of the bank loans; the effect of subsidies was stronger than that of bank loans. Girma et al [10] examines the impact of enterprise support on firm survival and growth in case of Irish manufacturing enterprises. In particular their study was special that in Ireland the public grants to enterprises have been used in addition to the improvement of domestic firms' performance also for attracting the foreign firms' production units to the country. They used traditional matching techniques in combination with difference-indifference analysis and showed that especially capital (but also other types of) grants had important impact on firm survival and job creation. The main finding of Ege [7] is that the Small Innovative Research grants in USA stimulate both sales and employment growth. These results are robust across several alternative regression models and different groups of control variables. The most important control variables were the firm's sales in the year of application and the firm's employment in the year of application. Sissoko [17] investigates the role of R&D subsidies on productivity of the French firms. He explores their role on the firm performance measures like employment, capital and R&D expenditures using difference-in-difference techniques. The results suggest that, on average, total factor productivity of the subsidized firms is higher of around 15% towards the end of the 3-years grant period relative to the matched control group. There is also little evidence about a role of R&D subsidies on employment, capital, R&D expenditures and credit constraints. The recent research of impact of subsidy was done by Criscuolo et al [6] in Great Britain. They analyzed the impact of expenditure on the Regional Selective Assistance program over a 20-year period. They had over 2.3 million observations before and after receiving government support. Using IV estimates they found positive program treatment effect on employment, investment and net entry but not on productivity. Their research suggests that government grants to smaller firms in economically disadvantaged areas of Great Britain can increase employment, but that grants to larger firms have no effect.

Moving on to the existing studies in Estonia there is empirical research analyzing only the impact of start-up grants on firms' efficiency. Lukason and Masso [13] analyzed the performance of 39 Estonian start-up firms that received financial aid from the state in the form of start-up grant during 2005-2008. The results indicated that while many firms could not meet their reported goals (in terms of turnover, profit and the number of jobs created) and more than half of the firms had tax arrears, the estimated labor taxes paid by these firms were much higher compared to the sum of the grant, thus indicating the positive net impact of grants on the state's fiscal position. Also Masso and Vildo [14] found that start-up grants had positive impact on job creation in second year after getting the grant, but for all viewed years concerning the sales growth. At the same time they concluded that start-up grants did not increase firm's survival chances.

3. Data and methodology

This paper employs unique data from Enterprise Estonia and Estonia Central Commerce Register. The firms are grouped according to the Estonian Classification of Economic Activities (EMTAK). The period covered is 2004 to 2010. In current study we exclude the firms which do not have EMTAK code or were from economic activities which did not receive the financial support (agriculture, forestry and fishing). We exclude big firms

with more than 3000 employees. The number of firms by year of receiving grant as follows 2004 – 45 firms, 2005 – 53, 2006 – 24, 2007 – 48, 2008 - 291 and 2009 – 75 firms. The number of firms that received any EAS grant is 536. The comparison group consists of 40275 enterprises and comes from Estonia Central Commerce Register.

In the paper are used two different methods. To analyze the impact of the financial support we applied the difference-in-differences (DID) estimator using a regression framework and a panel data framework.

DID has become more popular in the estimation the causal relationship. The idea behind DID is to compare the outcome in the case of one group of the firms that received the grant with the outcome in the case of another group that did not and then to compare their before-and-after levels [9], [12].

The impact is calculated using the difference between pre- and post- intervention mean outcomes for the treatment and comparison groups and then subtracting the two differences:

$$impact = \delta = \left(Y_{post}^{treat} - Y_{pre}^{treat}\right) - \left(Y_{post}^{compar} - Y_{pre}^{compar}\right) \tag{1}$$

The first difference controls for time-invariant factors. The second difference controls for time-varying factors that are the same in both treatment and comparison group. Thus, selection bias is eliminated due to differencing.

The regression of output can be summarized in the following simple regression:

$$Y_{i10} = \beta_0 + \alpha T_i + \sum_{k=1}^{m} \beta_k X_{ki} + u_i. \tag{2}$$

$Y_{i10} = Y_{i1} - Y_{i0}$, where Y_{i1} is firm i dependent variable in the evaluation year belongs to the comparison group and Y_{i0} is the same variable in the base year. The parameter T_i=1 if firm belongs to the treatment group, 0 if it is comparison group. The treatment group consists of firms that received grants during the study period and comparison group are other firms according to Estonia Central Commerce Register. X_{ki} stands for the other observable characteristics of firm i, u_i error term. As there is heteroscedasticity in all models analyzing the relationship between financial support and firm's outcome by economic activities we imply heteroscedasticity adjusted standard errors [11], [4].

The next method of the analysis is the panel data approach. This approach is used if the estimation effect has impact on the individuals in the different periods [12], [3]. The regression equation takes the form:

$$Y_{it} = \beta_0 + \gamma H_{it} + \beta_1 X_{1it} + u_{it} \tag{3}$$

where t denotes the time index and i the firm index. The parameter H_{it} has been defined as a dummy variable, where the variable takes 1 it the firm has obtained the grant before the evaluation year and 0 otherwise. X_{ki} stands for the other observable characteristics of firm i, u_{it} error term.

We use the fixed effects (FE) panel model in this paper. This choice is reasonable as our data consists of almost the all firms that received the grants. In that case we assume that the u_{it} may be correlated with some of the regressors in the model. Similarly, Hausman test shows that random effects model is redundant.

The model that provides the overall theoretical framework and estimating equation for this paper is derived from a Cobb-Douglas production function: $Q = AK^{\alpha}L^{\beta}$ where Q is output, K is capital and L is labor, which can be written in logarithmic intensive form as:

$$\ln(MT_{it}) = \beta_{0i} + \gamma H_{it} + \beta \ln L_{it} + u_{it} \tag{4}$$

Hereby t denotes the time index and i the firm index. The parameter MT_{it} stands for net sales as output, H_{it} represents the obtaining grant and L_{it} is the number of employees. Unfortunately, we have not obtained the data about capital so far, we could not apply firms' assets total into the model.

As the period 2004-2010 includes the both rapid economic growth and recession, we include to the model parabolic trend. We consider the fixed effects model by adding time effects to the model:

$$\ln(MT_{it}) = \beta_{0i} + \gamma H_{it} + \beta_1 \ln L_{it} + \beta_2 \cdot t + \beta_3 \cdot t^2 + u_{it} \tag{5}$$

Hereby the parameter t stands for time variable.

We used robust clustered standard errors to account for the possible within-group correlation. This is usual procedure for grouped data because the performance of firms within a country may be somehow correlated and it

is not possible to capture all of this correlation with available set of explanatory variables. Another reason for clustering arises from the inclusion of group level variables together with firm-level variables in the same regressions. The "cluster" adjusted standard error (as performed in programs such as Stata) is aimed at dealing with the within group correlation structure but does not impose homogeneity of the variances [15], [4].

4. Results and Discussion

Many types of subsidy have been used in Estonia to support enterprises. In this study, we concentrate on some of them that were received in EAS:

1. Start-up and development grants is to provide support for starting companies in investments related to starting and developing a business,
2. Research and Development (R&D) grant for creating the good products and services in cooperation with entrepreneurs and scientists,
3. Development of Knowledge and skills project grant is meant for projects aimed at developing entrepreneurship and increasing business knowledge and activity,
4. Technology investment grant for industrial enterprises,
5. Export grant is to promote the export activities.

Table 1 presents summary statistics of the received grants for the treatment group from Enterprise Estonia (EAS). It is evident that the number of firms that received development of knowledge and skills grants is greater among received firms. Also the sum of R&D grants is one of the biggest among other grants. It means that it is important measure of industrial policy to promote innovations and R&D activities.

Type of grant	count	mean	min	max	Std.dev	cv	sum
Development grant	24	11404	4470	12782	2428	0.213	273703
Start-up grant	93	3159	655	6391	910	0.288	293766
Export program	91	49372	1853	63912	20527	0.416	4492808
R&D grants	62	69664	3323	564117	124198	1.783	4319152
Development of knowledge and skills	246	5239	1598	65190	6959	1.328	1288873
Technology investment program for industrial enterprises	20	193987	30678	900510	221484	1.142	3879733
TOTAL	536	27142	655	900510	72528	2.672	1.45×10^7

Table1 Summary statistics for the firms that received grant, EUR

The presentation of our results is divided into two sections according to the method that is used for the analysis of impact of grant on firms' performance. Firstly, we used difference-in-difference approach (2). The results are presented in Table 2.

Variable	R&D grant		Development of knowledge and skills grant		Technology investment grant for industrial enterprises	
T	-5.98×10^5		1.45×10^6		1.53×10^6	
Significance of T	0.076		0.093		0.183	
Net sales 2008	-0.441 ***	-0.441***	-0.448 ***	-0.449 ***	-0.441 ***	-0.441***
Number of employees 2008	5.14×10^5 ***	3.29×10^4 ***	3.77×10^4 **	3.73×10^4 ***	3.29×10^4 ***	3.29×10^4 ***
R^2	0.475	0.569	0.519	0.519	0.569	0.569
Observations	24465		24615		24474	

Table 2 Difference-in-difference results. Dependent variable is net sales 2010, base year is 2008
*Note: ***, ** and * denote statistical significance at 1, 5 and 10% respectively. In all models constant and 16 dummy variables (by 17 economic activities) are included.*

An investigated outcome variable in this paper is net sales of firms. Explanatory variables are net sales and number of employees in the base year. Due the sample consists of the firms that received grants in the different years, we define evaluation and base year (2010 and 2008 respectively). We include to Table 2 only statistically significant results for R&D grant, development of knowledge and skills grant and technology investment grant for industrial enterprises. The indicator of received grant before evaluation period shows the negative impact on

net sales for R&D grant and positive for other grants. One can see that the impact of two grants, R&D grant and Development of knowledge and skills grant, is significant on 0.1 level.

During the period 2004-2010 firms received grants at different time and here are 16 different combinations of base year and evaluation year. DID method is applicable separately for each base year/evaluation year combination. Secondly, panel regression analysis used a cross-sectional database composed of 39484 firm observations. Table 3 shows the results according to received grants. We can see the positive impact of grant on the firms' performance almost in all models. It is significant at a 5% level for start-up grant, export grant, development of knowledge and skills grant and technology investment grant for industrial enterprises.

To sum up we can evaluate the impact of grant on Estonian firm performance indicator as follows:

1. Start-up grant increases net sales on average by 30%,
2. Export grant increases net sales on average by 11%,
3. Development of knowledge and skills grant increases net sales on average by 20%,
4. Technology investment grant for industrial enterprises increases net sales on average by 33%.

Variable	**Start-up grant**	**Development grant**	**Export grant**	**R&D grant**	**Development of knowledge and skills**	**Technology investment grant**
Impact of grant	0.301***	0.250	0.107***	0.106	0.204***	0.326***
Number of employees (logaritmic)	0.786***	0.786***	0.785***	0.785***	0.784***	0.785***
t	0.242***	0.242***	0.242***	0.242***	0.242***	0.242***
t^2	-0.034***	-0.034***	-0.034***	-0.034***	-0.034***	-0.034***
intercept	10.21***	10.21***	10.21***	10.21***	10.21***	10.21***
R^2 overall	0.524	0.524	0.529	0.525	0.531	0.525
Number of firms	39484	39417	39484	39451	39639	39413

Table 3 Panel regressions of different types of grants. Dependent variable is logarithmic net sales
*Note: ***, ** and * denote statistical significance at 1, 5 and 10% respectively.*

Variable	**Dummies for different grants**		**All grants with grant amount**	
	Coefficient	p-value	Coefficient	p-value
H	0.326	0.000	0.186	0.000
Development grant dummy	-0.076	0.649		
Start-up grant dummy	-0.025	0.810		
Export grant dummy	-0.220	0.002		
R&D grants dummy	-0.220	0.061		
Development of skills grant dummy	-0.123	0.063		
Grant amount			2.58×10^{-7}	0.216
Number of employees (logarithmic)	0.784	0.000	0.784	0.000
t	0.242	0.000	0.242	0.000
t^2	-0.034	0.000	-0.034	0.000
intercept	10.22	0.000	10.22	0.000
Sign. F-test (all grant dummies=0)	0.034			
R^2 overall	0.538		0.538	
Number of firms	39923			

Table 4 Panel regressions of all types of grants. Dependent variable is logarithmic net sales

Then we included into the model five dummies for six types of grants in the model (5). The results are showed in Table 4. The base type was the grant „Technology investment grant for industrial enterprises". The coefficient of the dummy variable H is now the same as in the model for "Technology investment grant" (Table 3). The coefficients of the other grants dummies are differential intercept coefficients and they tell us how much the impact of the other grants is different with respect to the base type. Using these differences we can calculate the values of the impact coefficients for all models in Table 4. To test whether categorization by grant types is relevant or not, we use the F-test on parameters restriction with null hypothesis that all grant dummies are jointly zero. P-value of the F-test was 0.034, so we can conclude that different grants have different impact to the net sales of firms.

Table 1 shows that various grants on average have different amounts and the observed differences may be due to this. To test the hypothesis we added into the model variable amounts of grants as are shown in Table 4. Since the contribution of the impact of grants on the net sale do not proceed in the same year, the impact of grants will be nonzero only in the evaluation year, if $H = 1$. It is seen that amounts of grants is not statistically significant (p-value 0.216). From the last table we can conclude that the impact of various grants is different on the net sales. The most impact can be seen for technology investment grant for industrial enterprises.

5. Conclusions

The paper contributes to the literature on evaluating the effectiveness of government grants given to enterprises in Estonia. Data of the firms are collected from Enterprise Estonia and Estonia Central Commerce Register from 2004 to 2010. The dependent variable is only the net sales in the paper. Econometric analyses show that these programs have positive impacts on firms' performance in case of the Estonian firms. Government grants have different aims and they are allocated in accordance with different criteria. Therefore, assessing the impact of grants on the effectiveness measures should be applied with different evaluation criteria.

Acknowledgements

This research was supported by the Ministry of Education and Research of the Republic of Estonia under Grant No. SF0140059s12.

References

[1] Almus, M.: Evaluating the Impact of Public Start-up Assistance – Results form an Econometric Approach, ZEW Discussion Paper No. 01-23, 2001.
[2] Bergström, F.: Capital Subsidies and the Performance of Firms, *Small Business Economics*, **14** (2000), 183-193.
[3] Bertrand, M., Duflo, E. and Mullainathan, S.: How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics* **119** (2004), 249-275.
[4] Cameron, A. C., and Trivedi, P. K.: *Microeconometrics Using Stata*. A Stata Press Publication, 2009.
[5] Crepon, B., and Duguet, E.: Bank loans, start-up subsidies and the survival of new firms: an econometric analysis at the entrepreneur level, Economics Working Paper Archive, Labor and Demography No 0411004, 2003.
[6] Criscuolo, C., Martin, M., Overman, H., and Van Reenen, J.: Causal Effects of an Industrial Policy. NBER Working Paper No. 17842, 2012.
[7] Ege, M.: How do grants influence firm performance? An econometric evaluation of the SBIR program at NIH. New Brunswick, New Jersey, 2009.
[8] Enterprise Estonia, http://www.eas.ee/en
[9] Garcia, M.: Micro-methods in evaluating governance interventions. Discussion Paper No.7, German Development Institute, 2011.
[10] Girma, S., Görg, H., and Strobl, E.: Government Grants, Plant Survival and Employment Growth: A Micro-Econometric Analysis", IZA Discussion Paper No. 838, 2003.
[11] Greene, W. H.: *Econometric Analysis.* 5th ed., Prentice Hall, 2002.
[12] Khandker, S. R., Koolwalt, G. B., and Samad, H. A.: *Handbook on Impact Evaluation: Quantitative Methods and Practices.* The World Bank, 2009.
[13] Lukason, O., and Masso, J.: Performance of selected Estonian firms financed with start-up grant: ability to follow plans and grant usage efficiency, Berlin: Berliner Wissenchafts-Verlag, Mattimar, 2010, 253 – 265.
[14] Masso, J., and Vildo, S.: The government grants to enterprises: theoretical arguments and the possibilities to measure their efficiency. Berlin-Tallinn: Berliner Wissenchafts-Verlag, Mattimar, 2006, 521 - 542.
[15] Petersen, M. A.: Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. *Review of Financial Studies*, **22** (2009)**,** 435-480.
[16] Sissoko, A.: R&D Subsidies and Firm-Level Productivity: Evidence from France. Discussion Paper, No.2, 2011.

Determination of mutually acceptable price of used manufacturing equipment

Simona Hašková[1], Pavel Kolář[2]

Abstract. This paper presents a general methodology for estimating the optimal acceptable price of used manufacturing equipment offered for sale because of the regular replacement of an old but still functional device with a new one. The optimality criterion for estimating the selling price is the maximization of net present value (NPV) of the production, from which the machine is phased out. The mutually acceptable price is the price that will improve the economic balance of the operation not only to the seller but to the buyer as well (in comparison with the acquisition and operation of new machinery in production).
Solution of this issue is primarily based on the assumption of the replacement with the same type of machine, of which the investment price and annual operating costs accruing during its physical life, is known.
The contribution states the original calculation procedure of the interval limits, in which the mutually advantageous price has to lie. The interpretation is demonstrated in a case study solved by the approach of "case-based reasoning."

Keywords: operating expenses, annual expenses, tax savings, physical life, economic life, PV costs, annuity factor, the equivalent of annual expenses.
JEL Classification: C58

AMS Classification: 91G50

1 Introduction

To the main results of the cognition process of the economic nature and its application to management theory belongs both "knowledge" (systems of management theory) and secondarily a rational "action" (methods of solving specific managerial tasks), which is derived from it.

The academics are mostly involved in the issue of knowledge within the basic research (inferring the knowledge from the observed data). They stand on the top of the imaginary pyramid of the process of cognition. They are followed by specialists who built the knowledge in the various management theories. At the bottom of this pyramid there are managers who draw on these theories that help them to solve specific tasks.

The working tool of academics is the generalization consisting of "trimming" the specifics and irrelevant details to describe the patterns of the widest class of the observed cases. The specialists built these patterns into their management theories in the way of adequate interpretation and they complement them by the essential specifics of the appropriate field. The intention of the specialist is to provide managers with guides (manuals) to solve their problems. However, in practice, managers rely on these guides generally only because they (in metaphorical terms) allow them often build only "panel houses" (acceptable solutions) and not "family villas" (optimal solution), for which they aspire. For the construction of "family villas" it is necessary, in addition to knowledge of management theory, the experience and a certain portion of the art (see [1]).

To consider the experience and the art in the solutions and to bring them closer to the optimal solution allows the approach commonly known among managers as "case-based reasoning" (see [6]). This approach does not delete the specifics of a particular task but rather uses them and respects them to the maximum extent, which allows attaining the desired unique solution. The outcome of this approach is a case study that solves a particular case, in which the results are not only presented but also adequately justified.

This contribution differs from the practice by the fact that there is not applied an approach of solving only one singular problem by "case-based reasoning" but that this approach is used for derivation of the method for solving certain classes of problems.

[1] VŠE Praha/Fakulta managementu JH, katedra managementu podnikatelské sféry, Jarošovská 1117/II, 377 01 Jindřichův Hradec, haskovas@post.cz.
[2] VŠE Praha/Fakulta managementu JH, katedra managementu podnikatelské sféry, Jarošovská 1117/II, 377 01 Jindřichův Hradec, kolar@fm.vse.cz.

The essential features of the specific tasks included under the title of this article can be specified as follows:

- The task can be successfully solved only in the context of problems associated with it. The context is:
  - the selection of the type of the equipment (machine),
  - time of its decommissioning,
  - the way of its subsequent use and what it is to be replaced with.
- The considered variants of machines differ from each other by their physical lifetime. Other relevant information on input, in addition to the purchase price, is the cost characteristics of the machines. Benefits from their use are given by the inclusion in the same manufacturing process and they do not depend on cost characteristics of individual machines.
- The question is whether the sale of the used machine at the time of its working life can improve the economic balance of its operation.

In the next part, the problem will be firstly formulated in general and then gradually solved. The interpretation of the method will be accompanied by examples of the case study of the problem of two plate bending machines (machines Standard and Prime) addressed in detail in [3].

2 The formulation of the problem in general

When estimating limits of the reasonable price of the used machine we need to take into account the aspects of both sides - the buyer and the seller. We expect that both of them respect the rule of maximization net present value (NPV) based on risk aversion (see [4]) during this transaction:

- The seller decides according to the production program, in which the machine is included. If the machine needs to be replaced, he deducts from the sale income P_1 the present value PV_1 of increases of current and future expenses that relate to the contemplated replacement. As "present" he considers the moment of the income from sale. For a reasonable price P_1 from his point of view applies $P_1 - PV_1 > 0$.
- The buyer compares $PV_2 + P_1$, where P_1 is the price of the machine and PV_2 is present value of expenses related to its use with PV costs associated with the realization of his alternative solution ("present" is the moment of payment for the machine). For a reasonable price P_1 from his point of view applies $PV > P_1 + PV_2$, or $0 > P_1 + PV_2 - PV$.
- As the purpose and expected time of further use of the used machine can be generally different for the buyer and the seller, this business transaction may be the game with non-zero sum, from which ultimately benefit both. In this case, the solution of systems of inequations $P_1 - PV_1 > 0$ and $0 > P_1 + PV_2 - PV$ is a non-empty interval of values P_1. Its boundaries are the limits of a mutually acceptable price P_1. A fair compromise will be then the price located in the middle of it.

We assume there are the following data available: the cost of the machine, its depreciation plan, the time course of operating costs by the expected standard machine use during its physical lifetime, the income tax rate and capital cost of the project, with which it is connected (the discount rate). The formulation of relations for the relevant calculations is required.

3 The general solution procedure

In accordance with the specifics of a given class of tasks (irrelevance of benefits) as it is stated above, in case of selection of the machine, the criterion of maximization of net present value (NPV) of its cash flow reduces to the criterion of minimizing present value (PV) of costs associated with the acquisition and operation. The first phase of calculations is the conversion of the acquisition and operating costs (taking into account the tax savings) to the annual expenditures (the general conversion algorithm is shown in the first column of the tables stated in part 4 below).

3.1 Equivalent of annual expenditure

When considering several variants of machines with unequal periods of physical lifetime, we can not judge this option by PV expenses (see [2]). In this case, the irrelevance assumption of benefit does not apply (PV benefits of the machine working longer time is greater). One possibility is to compare the PV(n) of expenditures of variants for the interval of *n* year's length, where at the end all compared machines would be removed simultaneously. This, however, happens if *n* is the lowest common multiple of the physical lifetime of compared variants. The interpretation of this idea is simple: Just imagine that the company always replaces the decommissioned machine by exactly the same type of machine producing over its lifetime the same flow of costs and benefits as the decommissioned one. In the moment when comes to the decommissioning of all machines at the same time, the PV(n) of benefits of all variants is the same (i.e. irrelevant). The individual variants differ only in the PV(n) costs. The lower PV(n) is the better.

However, there is a better option: Imagine that we calculated for the *n*-length period of the mentioned interval of the lowest common multiple the PV(n) of actual annual expenditures and ask what would be a constant annual expenditure, by which we would replace the actual annual expenses so that the resulting effect for the period (i.e. PV(n) costs) was in both cases the same. It is obvious that we basically ask about the amount of annuity payment of debt in the amount of PV(n) by *n*-year annuity. This "annuity payment" is called **the equivalent of annual expenses** and we mark it as ERV(n). Instead of PV(n) we can then evaluate the variants of the machines according to ERV(n). Again, the lower ERV(n) is the better. It is obvious that applies:

$$ERV(n) = \frac{PV(n)}{\sum_{i=1}^{n}(1+r)^{-i}} = \frac{PV(n)}{(r^{-1}\cdot(1-(1+r)^{-n})} = \frac{PV(n)}{\alpha(r,n)} \tag{1}$$

where *r* is the discount rate, α (r, n) *n*-year annuity factor and Σ is the symbol of summation over *i* from 1 to *n*.

We see that the ERV(n) is generally a function of length of the period *n*. However, in [3] it is proved that for any natural number *n* holds: If the subsequent expenditure in a year n + 1 has the value ERV(n) then ERV(n + 1) = ERV(n). Otherwise (i.e. if the subsequent expenditure is higher or lower than the previous ERV(n)) this expenditure is "pulling" the ERV(n + 1) toward itself, so ERV(n + 1) ≠ ERV(n).

The result of what was said above is the fact that for n = k · z, where k ≠ 0 is a natural number and z is the physical lifetime of the machine the following applies:

$$ERV(n) = ERV(k\cdot z) = ERV(z) = \frac{PV(z)}{\alpha(r,z)} \tag{2}$$

The calculation of the ERV is then quick and easy, not only because annuity factors are tabulated, but mainly due to equation (2) it is possible to get the result from the PV(n) expenses and from the annuity factor for the physical lifetime of the variant without a need to seek a common multiple.

3.2 Physical versus economic life

So far we assumed the operation of the machine through its whole physical lifetime and we considered its replacement after the end of this period when its physical breakdown threatened. This strategy is not optimal as follows from the following consideration:

A major expense is acquiring the machine in year 0 and the value ERV(0) approaches it. The new machine tends to have relatively low operating costs. They pull down ERV value at first. As the number of years of the machine operation grows, it wears out, which can result in growing operating costs in such a way that they pull ERV up. In the set {1, 2, ..., z} of years of the machine operation, where *z* is the time of its physical life, then for some i ∈ {1, 2, ..., z} exists ERV(i) = ERVE so that for every n ∈ {1, 2, ..., z} applies ERV(n) ≥ ERVE. The value ERVE = ERV(i) = minERV in the set {1, 2, ..., z} and the time *i* is **the economic lifetime** of the machine.

The interpretation of the concept of economic life is obvious. It is the time of machine operation when it is convenient to replace it with a new one. This results in achieving the minimum annual costs in the amount of ERVE for the machine operation including its acquisition.

3.3 The acceptable price of the used machine

Mutually acceptable price for the used machine is the price that will improve the economic balance of the operation not only to the seller but to the buyer as well (in comparison with the acquisition and operation of the new machine in its production). In the following part we denote the above defined ERVE of the seller as $ERVE_1$ and of the buyer as $ERVE_2$. Generally $ERVE_1 \neq ERVE_2$ because the buyer can use the machine in a different mode than the seller (a different timing of expenditures) or he uses it only for a shorter period than its economic life is. The case $ERVE_1 = ERVE_2$ is interesting if we ask whether it is economic to use in the production process the machine from "second hand" only.

Let us denote:

x .....a reasonable sale price of the used machine (in thousands $)

t ..... economic lifetime of the machine (in years)

k .....number of years of use before decommissioning and sale (age of sold machines)

v_j ..... expenditures for machine in j-th year of use,

n_j ..... machine operating costs in j-th year of use,

zc_k .....net book value of the decommissioned machine

r ..... discount rate,
T ..... income tax rate.

We assume that at the time of the machine decommissioning its net book value is depreciated and that the buyer of the used equipment depreciates the entire investment (purchase price) in the first year of use:

- For the requirement of a reasonable sale price of the used machines that should provide the seller with ERV of the previous operation not exceeding $ERVE_1$, we look at PV_1 flow of expenditures that involves capital expenditure and the sum of discounted operating expenditures, from which we deduct the amounts reducing costs in the form of tax relief from the net book value depreciation and the discounted net income from the sale of the machine (after taxation). Mathematically this problem can be formulated as:

$$PV_1 = v_0 + \sum_{j=1}^{k} \frac{v_j}{(1+r)^j} - \frac{T \cdot zc_k}{(1+r)^k} - x \cdot \frac{(1-T)}{(1+r)^k} \leq \sum_{j=1}^{k} \frac{ERVE_1}{(1+r)^j} \tag{3}$$

After adjustments we get:

$$x \geq (1+r)^k \cdot (v_0 - \sum_{j=1}^{k} \frac{ERVE_1 - v_j}{(1+r)^j}) / (1-T) - \frac{T \cdot zc_k}{1-T} \tag{4}$$

- When formulating the acceptable price for the buyer that should ensure ERV of the subsequent operation not exceeding $ERVE_2$ we come out from PV_2 flow of expenditures again, but this time from the perspective of the buyer. The investment expenditure is in this case the purchase price reduced of the discounted tax relief from its depreciation. The sum of discounted operating costs is added to it and is reduced of the tax relief resulting from it. In the subsequent mathematical formulation of the problem we get:

$$PV_2 = x \cdot \frac{1-T}{1+r} + (1-T) \cdot \sum_{j=1}^{t-k} \frac{n_{k+j}}{(1+r)^j} \leq \sum_{j=1}^{t-k} \frac{ERVE_2}{(1+r)^j} \tag{5}$$

After adjustments we get:

$$x \leq (1-T)^{-1} \cdot ERVE_2 \cdot \sum_{j=1}^{t-k} \frac{1}{(1+r)^{j-1}} - \sum_{j=1}^{t-k} \frac{n_{k+j}}{(1+r)^{j-1}} \tag{6}$$

If the sets of inequalities (4) and (6) have a non-empty intersection, then this intersection defines the interval of mutually acceptable prices of the used machine. At a price in the middle of this interval both sides from this transaction get the same benefit.

4 The demonstration of calculations in the case study[3]

This part is based on a case study addressed in detail in [3]. The two machines (Standard and Prime) with uneven physical lifetime (6 and 9 years) are compared. The machines do the same thing in a different way (based on different principles). The cost characteristics of these machines are listed in the fifth line of the following tables, in which the detailed procedure of the first phase of the solution is recorded (conversion of purchase costs, operating costs and tax relief to annual expenditures). Expected tax rate is T = 34 % and the discount rate (cost of capital) is r = 10 %. The amounts are in thousands of $:

Machine Standard / year	**0**	**1**	**2**	**3**	**4**	**5**	**6**
1. Investment	35						
2. Depreciation	11.67	15.55	5.183	2.593			
3. Net book value		7.777	2.593				
4. Tax relief from depreciation[4]	3.97	5.29	1.76	0.88			
5. Operating costs		20	23	30	40	55	80

[3] For the purpose of explaining the method we are using the appropriately adjusted data suitable for demonstration of calculation procedures (tax rate and accelerated depreciations valid in the USA) – for details see [2] and [3].

[4] Tax relief from depreciation is a multiple of the annual depreciation and the expected tax rate.

6. Tax relief from operating costs[5]		6.8	7.82	10.2	13.6	18.7	27.2
7. Expenditures (1+5-4-6)	**31.03**	**7.91**	**13.42**	**18.92**	**26.4**	**36.3**	**52.8**

Table 1 Conversion of purchase costs, operating costs and tax relief to the annual expenditures

Machine PRIM / year	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1. Investment	50									
2. Depreciation	16.7	22.22	7.4	3.7						
3. Net book value		11.11	3.7							
4. Tax relief from depreciation	5.67	7.56	2.5	1.2						
5. Operating costs		20	20.5	22	25	30	37	46	60	80
6. Tax relief from operating costs		6.8	6.97	7.5	8.5	10.2	12.5	15.6	20.4	27.2
7. Expenditures (1+5-4-6)	**44.3**	**5.6**	**11**	**13.3**	**16.5**	**19.8**	**24.4**	**30.3**	**39.6**	**52.8**

Table 2 Conversion of purchase costs, operating costs and tax relief to the annual expenditures

From the line 7 according to the equation (2) for the machine Standard we get (see Table 1):

$$ERV(z) = PV(z) / \alpha(r,z) = ERV(6) = PV(6) / \alpha(0.1;6) = 133.86 / 4.355 = 30.74 \tag{7}$$

For the machine Prim we get (see Table 2):

$$ERV(z) = PV(z) / \alpha(r,z) = ERV(9) = PV(9) / \alpha(0.1;9) = 162.29 / 5.759 = 28.18 \tag{8}$$

We see that by replacement based on the physical lifetime the machine Prim is preferable.

The following two figures show the course the function ERV(s), $s \in \{1, 2, 3, 4, 5, 6\}$ of the machine Standard and ERV(p), $p \in \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ of the machine Prim. In the first two years tax savings from depreciation of the net book value of decommissioned machine is taken into account in the calculations ERV. The figures show that the machine Standard has only half physical lifetime compared to its economic lifetime (3 years) with $ERVE_S = 25.54$. The machine Prim´s economic lifetime is 5 years with $ERVE_P = 24.29$. Even on the basis of economic lifetime the machine Prim is better.

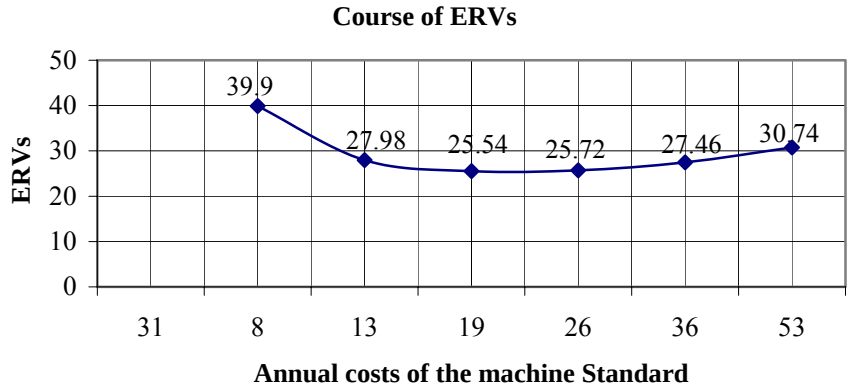


Figure 1 Course of the function ERV(s)

[5] Tax relief from operating costs is a multiple of the annual operating costs and the expected tax rate.

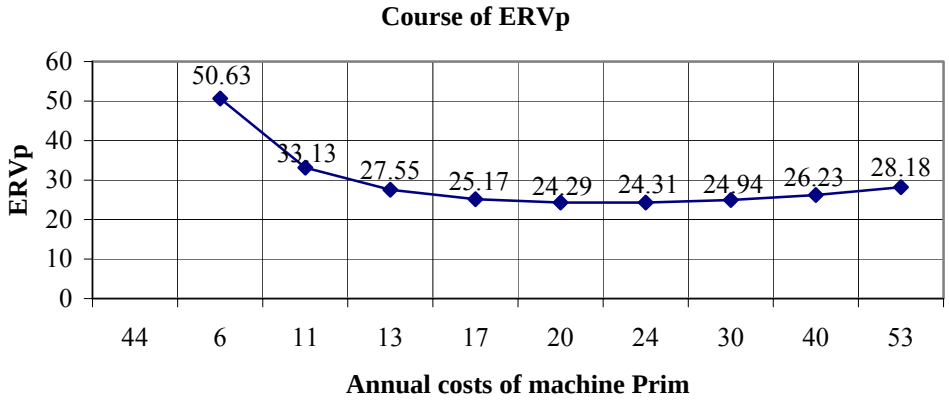


Figure 2 Course of the function ERV(p)

The case study addressed in [3], in addition to other things should have considered the possibility of repeated acquiring only used machine (i.e. the variant $ERVE_1 = ERVE_2$). The calculations results of inequalities (4) and (6) are summarized in the following table. Due to the fact that all the intervals in the table 3 are empty, we can conclude that this plan is not viable.

Age of the offered machine	Model PRIM	Model STANDARD
1 year	$40 \leq x \leq 38.8$	$21.04 \leq x \leq 20.53$
2 years	$28.12 \leq x \leq 27.08$	$7.77 \leq x \leq 7.57$
3 years	$16.4 \leq x \leq 15.65$	
4 years	$6.22 \leq x \leq 5.92$	

Table 3 Calculated intervals of mutually acceptable price

5 Conclusion

The paper presents and justifies in detail the original method of estimation of the interval limits of mutually favorable prices of the used manufacturing equipment. The method is based on "case-based reasoning"; the profitability criterion is NPV of benefits. Within the limited range of the article there could be presented only the basics of the method, at which possible extensions further build. One of them is the possibility of replacement of the existing machine with the new one and better one, which could be available in the near future. The uncertainty associated with it draws in the game various possible scenarios of solutions and requires a transition from the criterion NPV to criterion E[NPV] including all the problems associated with it (for more details see [3] and [4]). This, however, is beyond the scope of this paper.

References

[1] Ackoff, R. L.: *The Art of Problem Solving.* John Wiley&Son 1987.
[2] Brealey, R. A., and Myers, S. C.: *Teorie a praxe firemních financí*. Computer Press, Praha 2000.
[3] Hašková, S.: *Základy teorie a praxe manažerských propočtů*. Diplomová práce FM VŠE, 2010.
[4] Hašková, S., and Kolář, P.: A contribution to a mathematic modeling of the behavior of private and institutionalized investors as observed in practice. In: *Proceedings of the 28th International Conference on Mathematical Methods in Economics 2010*. University of South Bohemia in České Budějovice, Faculty of Economics, 2010, 214-219.
[5] Johnson, N.H., and Solomon, B.D.: A Net-Present Value Analysis for a Wind Turbine Purchase at a Small US College. In: *Energies 2010, www.mdpi.com/journal/energies*, Open Access, 943-959.
[6] Thagard, P.: *Úvod do kognitivní vědy, Mysl a myšlení*. Nakladatelství Portál, s.r.o. 2001.

Different approaches to dynamic conditional correlation modelling: the case of European currencies

Radek Hendrych[1]

Abstract. The analysis of time varying conditional correlation structures seems to be a significantly important part of multivariate time series modelling, particularly from the (practical) financial or economic point of view. In 2002, Robert Engle published an innovative concept in the framework of this issue. A simple class of multivariate autoregressive conditional heteroskedasticity models, the so-called dynamic conditional correlation models were introduced. Thereafter, these techniques have been examined and adjusted in many different theoretical or empirical ways. In the contribution, several various approaches to modelling the dynamic conditional correlations originally based on Engle's idea are reviewed and discussed. Some of their pros and cons are mentioned and demonstrated. Finally, the comparison of their performance is shown in the study of the portfolio of the European currencies and their correlation links. All the relevant procedures are implemented in the statistical software R.

Keywords: multivariate financial time series, conditional covariance, dynamic correlations, DCC, European currencies.

JEL classification: C32
AMS classification: 91B84

1 Introduction

Consider a stochastic vector process $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ of the dimension $(n \times 1)$. Denote $\mathcal{F}_{t-1}$ the information set (σ-algebra) generated by observed multivariate time series $\{\mathbf{X_t}\}$ up to and including time $t-1$. Let $\boldsymbol{\theta}$ be a finite vector of (real) parameters.

Assume the following model

$$\mathbf{X}_t = \boldsymbol{\mu}_t(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t = \mathbf{H}_t(\boldsymbol{\theta})^{1/2}\mathbf{Z}_t, \tag{1}$$

where $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ is the $(n \times 1)$ conditional mean vector of $\mathbf{X}_t$ and $\mathbf{H}_t(\boldsymbol{\theta})$ is the $(n \times n)$ conditional covariance matrix of $\mathbf{X}_t$. Furthermore, one supposes that $\{\mathbf{Z}_t\}$ is an $(n \times 1)$ i.i.d. stochastic vector process independent of $\{\mathbf{X}_t\}$ such that it has following first two moments: $\mathsf{E}(\mathbf{Z}_t) = \mathbf{0}$ and $\mathsf{var}(\mathbf{Z}_t) = \mathbf{I}_n$, where $\mathbf{I}_n$ is the $(n \times n)$ identity matrix.

The presented structure can be easily verified:

$$\mathsf{E}(\mathbf{X}_t|\mathcal{F}_{t-1}) = \boldsymbol{\mu}_t(\boldsymbol{\theta}) + \mathbf{H}_t(\boldsymbol{\theta})^{1/2}\mathsf{E}(\mathbf{Z}_t|\mathcal{F}_{t-1}) = \boldsymbol{\mu}_t(\boldsymbol{\theta}), \tag{2}$$

$$\mathsf{var}(\mathbf{X}_t|\mathcal{F}_{t-1}) = \mathsf{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top|\mathcal{F}_{t-1}) = \mathbf{H}_t(\boldsymbol{\theta})^{1/2}\mathsf{E}(\mathbf{Z}_t\mathbf{Z}_t^\top|\mathcal{F}_{t-1})(\mathbf{H}_t(\boldsymbol{\theta})^{1/2})^\top = \mathbf{H}_t(\boldsymbol{\theta}). \tag{3}$$

Thus, it is evident that $\mathbf{H}_t(\boldsymbol{\theta})^{1/2}$ is any $(n \times n)$ positive definite matrix such that $\mathbf{H}_t(\boldsymbol{\theta})$ is the conditional covariance matrix of $\mathbf{X}_t$, e.g. $\mathbf{H}_t(\boldsymbol{\theta})^{1/2}$ may be obtained by the Cholesky decomposition of $\mathbf{H}_t(\boldsymbol{\theta})$. Both $\boldsymbol{\mu}_t$ and $\mathbf{H}_t$ depend on the (unknown) parameter vector $\boldsymbol{\theta}$, which can be (in most cases) split into two disjoint parts, one for $\boldsymbol{\mu}_t$ and one for $\mathbf{H}_t$. The conditional mean vector is obviously specified as a linear model for the level of $\mathbf{X}_t$, e.g. VAR or VARMA. In the following section, several various specifications of the conditional covariance matrix $\mathbf{H}_t$ are reviewed.

[1]Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Prague 8, Czech Republic, e-mail: hendrych@karlin.mff.cuni.cz

2 Conditional covariance modelling

As stated above, the main objective is to capture the time varying behavior of the conditional covariance matrix $\mathbf{H}_t$ (for convenience the vector of parameters $\boldsymbol{\theta}$ is left out in the notation).

Bollerslev's *constant conditional correlation* (CCC) model (see [3]) decomposes the matrix $\mathbf{H}_t$ as

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R} \mathbf{D}_t, \tag{4}$$

where $\mathbf{D}_t$ is a diagonal matrix of time varying volatilities $\sqrt{h_{ii,t}}$, $i = 1, \ldots, n$, and $\mathbf{R}$ is an $(n \times n)$ constant conditional correlation matrix, i.e. a positive definite matrix with ones on its diagonal. The matrix $\mathbf{R}$ is usually estimated by the sample correlation matrix of standardized errors $\boldsymbol{\gamma}_t(\boldsymbol{\phi}) = \mathbf{D}_t^{-1}(\boldsymbol{\phi})\mathbf{X}_t$. The finite parameter vector $\boldsymbol{\phi}$ contains only relevant elements of the parameter vector $\boldsymbol{\theta}$. The diagonal elements of $\mathbf{D}_t$ can be modelled by usual univariate techniques for the conditional variance, e.g. by the univariate GARCH$(1, 1)$ model. However, the assumption that the conditional correlations are constant may seem unrealistic and be too restrictive.

Engle and Sheppard [6] offer an extension of the model (4) in a natural way to the more general case of *dynamic conditional correlations* (DCC) which are defined as

$$\begin{aligned} \mathbf{H}_t &= \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t, & (5)\\ \mathbf{R}_t &= \mathsf{diag}\{\mathbf{Q}_t\}^{-1/2} \mathbf{Q}_t \mathsf{diag}\{\mathbf{Q}_t\}^{-1/2}, & (6)\\ \mathbf{Q}_t &= (1 - \alpha - \beta)\mathbf{S} + \alpha \boldsymbol{\gamma}_{t-1} \boldsymbol{\gamma}_{t-1}^\top + \beta \mathbf{Q}_{t-1}, & (7) \end{aligned}$$

where $\mathbf{R}_t = \mathbf{R}_t(\boldsymbol{\phi}, \alpha, \beta, \mathbf{S})$ is a matrix of time varying conditional correlations with the unit diagonal elements, α and β are scalars and $\mathbf{S}$ is a parameter matrix. Specify that $\mathsf{diag}\{\mathbf{Q}_t\}$ is a diagonal matrix with $q_{t,11}, \ldots, q_{t,nn}$ on its diagonal, where $q_{t,11}, \ldots, q_{t,nn}$ are diagonal elements of the matrix $\mathbf{Q}_t$.

If $\mathbf{Q}_t$ is positive definite, $\mathbf{R}_t$ is also positive definite with unit diagonal elements. To ensure that $\mathbf{Q}_t$ is positive definite, it is sufficient to suppose that $\alpha \geq 0$, $\beta \geq 0$, $\alpha + \beta < 1$ and $\mathbf{S}$ is a positive definite matrix, see [6]. It is also frequent to assume that $s_{ii} = 1$, $i = 1, \ldots, n$, in order to guarantee the unique specification of $(\alpha, \beta, \mathbf{S})$. From (3) and (5), one can easily see that

$$\mathsf{var}(\boldsymbol{\gamma}_t | \mathcal{F}_{t-1}) = \mathbf{D}_t^{-1} \mathsf{var}(\mathbf{X}_t | \mathcal{F}_{t-1}) \mathbf{D}_t^{-1} = \mathbf{D}_t^{-1} \mathbf{H}_t \mathbf{D}_t^{-1} = \mathbf{R}_t. \tag{8}$$

With respect to the preceding assumptions, the DCC model defined by (5)-(7) contains $\frac{1}{2}n(n-1) + 2$ unknown parameters in addition to the parameters in $\boldsymbol{\phi}$. To eliminate complicated (quasi-)maximum likelihood estimation of all elements of the matrix $\mathbf{S}$ with many constraints, Engle [5] provides the so-called *correlation targeting*, i.e. $\mathbf{S}$ is substituted by the moment estimator $\hat{\mathbf{S}} = \frac{1}{T} \sum_{t=1}^{T} \hat{\boldsymbol{\gamma}}_t \hat{\boldsymbol{\gamma}}_t^\top$, $\hat{\boldsymbol{\gamma}}_t = \mathbf{D}_t^{-1}(\hat{\boldsymbol{\phi}}) \mathbf{X}_t$.

However, replacing $\mathbf{S}$ by $\hat{\mathbf{S}}$, i.e. by the sample second moment of $\boldsymbol{\gamma}_t$, is not suitable estimation device. The matrix $\hat{\mathbf{S}}$ is a biased and inconsistent estimator of $\mathbf{S}$, see [1] for more details. This can be easily shown: Suppose that $\alpha + \beta < 1$ and that $\mathsf{E}(\mathbf{Q}_t)$ and $\mathsf{E}(\boldsymbol{\gamma}_t \boldsymbol{\gamma}_t^\top)$ are independent of t. Thus, by applying the expectation operator on both sides of (7), the following equality is obtained

$$\mathbf{S} = \frac{1 - \beta}{1 - \alpha - \beta} \mathsf{E}(\mathbf{Q}_t) - \frac{\alpha}{1 - \alpha - \beta} \mathsf{E}(\boldsymbol{\gamma}_t \boldsymbol{\gamma}_t^\top). \tag{9}$$

Furthermore, it holds that $\mathsf{E}(\boldsymbol{\gamma}_t \boldsymbol{\gamma}_t^\top) = \mathsf{E}[\mathsf{E}(\boldsymbol{\gamma}_t \boldsymbol{\gamma}_t^\top | \mathcal{F}_{t-1})] = \mathsf{E}(\mathbf{R}_t) = \mathsf{E}(\mathsf{diag}\{\mathbf{Q}_t\}^{-1/2} \mathbf{Q}_t \mathsf{diag}\{\mathbf{Q}_t\}^{-1/2}) \neq \mathsf{E}(\mathbf{Q}_t)$, i.e. generally $\mathbf{S} \neq \mathsf{E}(\boldsymbol{\gamma}_t \boldsymbol{\gamma}_t^\top)$, apart from the case of constant conditional correlations.

Aielli [1] proposes a *corrected dynamic conditional correlation* (cDCC) model, namely the equality for $\mathbf{Q}_t$, i.e. (7), takes a different form:

$$\mathbf{Q}_t = (1 - \alpha - \beta)\mathbf{S}^* + \alpha \mathsf{diag}\{\mathbf{Q}_{t-1}\}^{1/2} \boldsymbol{\gamma}_{t-1} \boldsymbol{\gamma}_{t-1}^\top \mathsf{diag}\{\mathbf{Q}_{t-1}\}^{1/2} + \beta \mathbf{Q}_{t-1}. \tag{10}$$

Note that $\boldsymbol{\gamma}_t^* = \mathsf{diag}\{\mathbf{Q}_t\}^{1/2} \boldsymbol{\gamma}_t$ does not depend on $\mathbf{S}^*$ since the diagonal elements of $\mathbf{Q}_t$ only depend on the diagonal elements of $\mathbf{S}^*$ which are all equal to one, see the assumptions above. The matrix $\mathbf{S}^*$ in (10) can be consistently estimated by the sample second moment of $\boldsymbol{\gamma}_t^*$ due to two following facts. Firstly, from (6) and (8) it is clear that

$$\mathsf{var}(\boldsymbol{\gamma}_t^* | \mathcal{F}_{t-1}) = \mathsf{diag}\{\mathbf{Q}_t\}^{1/2} \mathsf{var}(\boldsymbol{\gamma}_t | \mathcal{F}_{t-1}) \mathsf{diag}\{\mathbf{Q}_t\}^{1/2} = \mathsf{diag}\{\mathbf{Q}_t\}^{1/2} \mathbf{R}_t \mathsf{diag}\{\mathbf{Q}_t\}^{1/2} = \mathbf{Q}_t. \tag{11}$$

Secondly, taking into account the similar assumptions, the same approach as in (9) and the fact that

$$\mathsf{E}(\mathbf{Q}_t) = \mathsf{E}(\mathsf{diag}\{\mathbf{Q}_t\}^{1/2}\mathbf{R}_t\mathsf{diag}\{\mathbf{Q}_t\}^{1/2}) = \mathsf{E}[\mathsf{diag}\{\mathbf{Q}_t\}^{1/2}\mathsf{E}(\boldsymbol{\gamma}_t\boldsymbol{\gamma}_t^\top|\mathcal{F}_{t-1})\mathsf{diag}\{\mathbf{Q}_t\}^{1/2}] = \mathsf{E}(\boldsymbol{\gamma}_t^*\boldsymbol{\gamma}_t^{*\top}), \quad (12)$$

the equality $\mathbf{S} = \mathsf{E}(\boldsymbol{\gamma}_t^*\boldsymbol{\gamma}_t^{*\top})$ holds.

The main difference in the correlation targeting in the models (7) and (10), i.e. substituting $\mathbf{S}$ and $\mathbf{S}^*$ by the sample second moment of $\boldsymbol{\gamma}_t$ and $\boldsymbol{\gamma}_t^*$, respectively, is that the matrix $\hat{\mathbf{S}}^*$ depends on the parameters (α, β) of the conditional correlation matrix, whereas $\hat{\mathbf{S}}$ does not.

Estimation of the previous (dynamic) conditional correlation models can be formulated as a maximum likelihood problem once a specific distributional assumption is made for the data. Obviously, it is supposed that the data are multivariate normal with the given mean and covariance structure. Fortunately, the considered estimator is a quasi-maximum likelihood, in the sense that it will be consistent but inefficient, if the mean and covariance assumptions are correctly specified even if other distributional assumptions are incorrect. See [1] or [4] for more information and references.

Thus, the log likelihood function for $\mathbf{X}_1, \ldots, \mathbf{X}_T$ for the model (5)-(7) can be written as

$$L(\boldsymbol{\phi}, \alpha, \beta, \mathbf{S}) = -\frac{1}{2}\sum_{t=1}^{T}(n\log(2\pi) + \log|\mathbf{H}_t| + \mathbf{X}_t'\mathbf{H}_t^{-1}\mathbf{X}_t) = L_V(\boldsymbol{\phi}) + L_C(\boldsymbol{\phi}, \alpha, \beta, \mathbf{S}), \quad (13)$$

where

$$L_V(\boldsymbol{\phi}) = -\frac{1}{2}\sum_{t=1}^{T}\left(n\log(2\pi) + \log|\mathbf{D}_t|^2 + \mathbf{X}_t^\top\mathbf{D}_t^{-2}\mathbf{X}_t\right), \quad (14)$$

$$L_C(\boldsymbol{\phi}, \alpha, \beta, \mathbf{S}) = -\frac{1}{2}\sum_{t=1}^{T}\left(\log|\mathbf{R}_t| + \boldsymbol{\gamma}_t^\top\mathbf{R}_t^{-1}\boldsymbol{\gamma}_t - \boldsymbol{\gamma}_t^\top\boldsymbol{\gamma}_t\right). \quad (15)$$

To split the function L into the sum of two parts L_V and L_C, one might use that $|\mathbf{H}_t| = |\mathbf{R}_t| \cdot |\mathbf{D}_t|^2$ and $\mathbf{H}_t^{-1} = \mathbf{D}_t^{-2} + \mathbf{D}_t^{-1}(\mathbf{R}_t^{-1} - \mathbf{I}_n)\mathbf{D}_t^{-1}$. Then, the estimation procedure is frequently done in two steps due to computational efficiency. Firstly, the parameters $\boldsymbol{\phi}$ of the time varying volatilities are estimated by maximizing $L_V(\boldsymbol{\phi})$. Secondly, the estimator $\hat{\boldsymbol{\phi}}$ of $\boldsymbol{\phi}$ from the preceding step and $\hat{\mathbf{S}}$ obtained by the correlation targeting (see above) are used in maximizing $L_C(\hat{\boldsymbol{\phi}}, \alpha, \beta, \hat{\mathbf{S}})$ to estimate the parameters (α, β).

Engle [4] refers to observed general downward bias of $(\hat{\alpha}, \hat{\beta})$ in the model (5)-(7) in the case of the second step of previous estimation. A simple adjustment with regard to this fact has been discovered, see [1]. The method is based on subsets of observations, i.e. on all combinations of pairs of elements $\{\mathbf{X}_t\}$. In particular, the composite likelihood function replacing L_C is proposed in the form

$$cL_C(\boldsymbol{\phi}, \alpha, \beta, \mathbf{S}) = \frac{1}{P}\sum_{p=1}^{P} L_{C,p}(\boldsymbol{\phi}, \alpha, \beta, \mathbf{S}), \quad (16)$$

where $L_{C,p}(\boldsymbol{\phi}, \alpha, \beta, \mathbf{S})$, $p = 1, \ldots, P = \frac{1}{2}n(n-1)$, denotes the bivariate (quasi-)log likelihood of the DCC submodel, i.e. the same as L_C in (15), defined only for two elements (one pair) from $\{\mathbf{X}_t\}$ with respect to the related adjustments of $\boldsymbol{\gamma}_t$, $\hat{\mathbf{S}}$, $\mathbf{D}_t$, $\mathbf{Q}_t$ and $\mathbf{R}_t$. Therefore, the estimators of (α, β) are obtained by maximizing (16). The method is computationally simple and does not require to invert large dimensional matrices. On the other hand, the composite likelihood estimators are less efficient than full maximum likelihood estimators.

In both previous cases, one can work also with the model based on the equation (10). It is sufficient to consider $\hat{\mathbf{S}}^*$, i.e. the sample second moment of $\boldsymbol{\gamma}_t^*$, instead of $\hat{\mathbf{S}}$ in the all preceding considerations. For other important characteristics of the models, e.g. their asymptotic distribution or the consistency of the estimators, see [1], [4] or [6].

3 EU currencies

To examine the empirical performance of the previously mentioned approaches to conditional correlation modelling, the exchange rates of the selected EU currencies are analyzed. In the EU27, 17 member

countries use the Euro, other 3 states (Denmark, Latvia and Lithuania) are members of the ERM II framework (the European Exchange Rate Mechanism II), i.e. the national currencies are allowed to fluctuate around their assigned value with respect to limiting bounds, and the Bulgarian Lev is pegged with the Euro. For these reasons, a portfolio of six remaining EU currencies is considered, i.e. the Czech crown (CZK), the British pound sterling (GBP), the Hungarian forint (HUF), the Polish zloty (PLN), the Romanian leu (RON) and the Swedish krona (SEK).

Particularly, logarithmic returns of the bilateral exchange rates from 2 January 2007 to 27 April 2012 (1365 observations) with the Euro as the denominator are considered, see Table 1. The data are available on the web pages of the European Central Bank.

	CZK	GBP	HUF	PLN	RON	SEK
mean	-0.00007	0.00014	0.00010	0.00006	0.00019	-0.00001
median	-0.00007	0.00012	-0.00024	-0.00014	0.00000	0.00004
maximum	0.03165	0.03461	0.05069	0.04164	0.02740	0.02784
minimum	-0.03274	-0.02657	-0.03389	-0.03680	-0.01992	-0.02260
std. dev.	0.00478	0.00601	0.00763	0.00721	0.00462	0.00497
skewness	0.20218	0.30655	0.42056	0.30802	0.54616	0.31526
kurtosis	8.49754	6.49258	7.80556	8.05110	7.37830	6.05079

Table 1: The basic characteristics of logarithmic returns of the selected exchange rates.

First of all, it is necessary to choose a suitable model for the conditional mean. Here, the VAR(3) model is used to catch the level of the data. This model is chosen with respect to results of the multivariate Ljung-Box test, several information criteria, the impulse response and root analysis.

Then, the conditional covariances are investigated. The performances of the CCC model, DCC model with MLE, DCC model with composite MLE, cDCC model with MLE and cDCC model with composite MLE are compared. Note that time varying volatilities, i.e. the first step of the estimation procedure, are represented by the standard EGARCH(1,1,1) model:

$$\log(h_{ii,t}) = \omega_i + \beta_i \log(h_{ii,t-1}) + \alpha_i \left| \frac{X_{i,t-1}}{\sqrt{h_{ii,t-1}}} \right| + \gamma_i \frac{X_{i,t-1}}{\sqrt{h_{ii,t-1}}}, \quad i = 1, \ldots, n. \tag{17}$$

Such a model is examined in several ways, e.g. the Ljung-Box and ARCH-LM tests of standardized residuals and squared standardized residuals for each of the univariate series. It is shown to be suitable in this way.

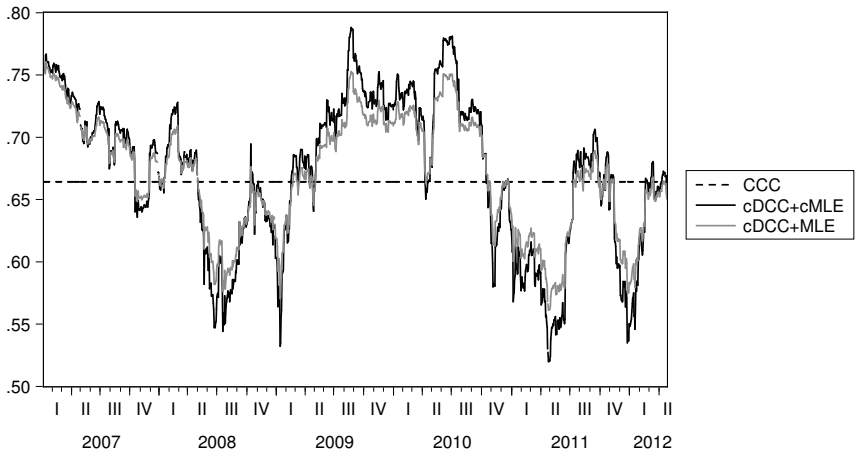


Figure 1: The various models of conditional correlations, the case of HUF/EUR vs. PLN/EUR.

The estimated parameters of conditional correlation modelling are surveyed in Table 2. The statistics of the multivariate Ljung-Box test with 7 lags are computed, i.e. the test of the null hypothesis of uncorrelated residual series. The corresponding p-values have mainly information value because the appropriate (asymptotic) distribution of the Ljung-Box statistic is not well known in these cases.

Hence, it is obvious that the estimators based on the DCC or cDCC models are quite similar. A certain difference can be observed between the estimation methods, namely MLE vs. composite MLE. The visual comparison of various models of conditional correlations can be seen in Figure 1 for the case HUF/EUR vs. PLN/EUR. The estimated correlation curve for the DCC with MLE (or with composite MLE) is not drawn due to its difficult distinguishability from the cDCC with MLE (or with composite MLE) in this scale, compare with Table 2.

	$\hat{\alpha}$	$\hat{\beta}$	$Q(7)$	(p-value)
CCC	-	-	245.942	(0.596)
DCC+MLE	0.01098	0.97833	252.878	(0.473)
cDCC+MLE	0.00945	0.98190	252.820	(0.474)
DCC+cMLE	0.01634	0.97432	254.689	(0.441)
cDCC+cMLE	0.01449	0.97783	254.415	(0.446)

Table 2: The estimators of (α, β) ($Q(7)$ refers to the multivariate Ljung-Box statistic with 7 lags).

For performance measures, the following regression-based tests are calculated on portfolio returns, $\mathbf{w}_t\mathbf{X}_t$, where $\mathbf{w}_t$ is a vector of portfolio weights. Note that the conditional variance of $\mathbf{w}_t\mathbf{X}_t$ is $\mathbf{w}_t^\top\mathbf{H}_t\mathbf{w}_t$. First, the Engle-Colacito (EC) regression is defined as $\{(\mathbf{w}_t\mathbf{X}_t)^2/(\mathbf{w}_t^\top\hat{\mathbf{H}}_t\mathbf{w}_t)\} - 1 = \lambda + \xi_t$, where ξ_t is an error term. The null hypothesis $\lambda = 0$ is verified. Point out that an HAC robust estimator of the standard deviation of ξ_t is required here. Second, the LM test of ARCH effects is based on the property that the series $\{(\mathbf{w}_t\mathbf{X}_t)^2/(\mathbf{w}_t^\top\mathbf{H}_t\mathbf{w}_t)\}$ does not exhibit serial correlation. The null hypothesis that $\{(\mathbf{w}_t\mathbf{X}_t)^2/(\mathbf{w}_t^\top\hat{\mathbf{H}}_t\mathbf{w}_t)\}$ is serially uncorrelated is tested (five lags are used). See [1] for more details.

Two types of portfolio weights are considered: the *equally weighted portfolio* (EWP), i.e. $\mathbf{w}_t = \mathbf{1}/n$, $\mathbf{1}$ is the $(n \times 1)$ vector of ones, and the *minimum variance portfolio* (MVP), i.e. $\mathbf{w}_t = (\mathbf{H}_t^{-1}\mathbf{1})/(\mathbf{1}^\top\mathbf{H}_t^{-1}\mathbf{1})$. The results are summarized in Tables 3 and 4 (σ denotes the standard deviation of the portfolio returns).

EWP	σ_{EWP}	EC-stat.	(p-value)	ARCH-LM(5)	(p-value)
CCC	3.6812E-03	-0.328	(0.743)	5.864	(0.320)
DCC+MLE	3.6812E-03	-0.611	(0.542)	3.115	(0.682)
cDCC+MLE	3.6812E-03	-0.877	(0.380)	3.238	(0.663)
DCC+cMLE	3.6812E-03	-0.621	(0.534)	2.463	(0.782)
cDCC+cMLE	3.6812E-03	-0.961	(0.337)	2.607	(0.760)

Table 3: The performance of the equally weighted portfolio.

MVP	σ_{MVP}	EC-stat.	(p-value)	ARCH-LM(5)	(p-value)
CCC	2.4816E-03	1.018	(0.309)	0.210	(0.999)
DCC+MLE	2.4447E-03	0.981	(0.327)	0.471	(0.993)
cDCC+MLE	2.4447E-03	0.820	(0.412)	0.458	(0.994)
DCC+cMLE	2.4417E-03	1.293	(0.196)	0.848	(0.974)
cDCC+cMLE	2.4414E-03	1.065	(0.287)	0.671	(0.985)

Table 4: The performance of the minimum variance portfolio.

From the practical point of view, the CCC model can give an idea of an average level of conditional correlations, see Figure 1. In Table 5, there is the estimator $\hat{\mathbf{R}}$ of the conditional correlation matrix $\mathbf{R}$. For instance, the British pound (GBP) does not have strong correlation links with other considered currencies. Further, the Hungarian forint (HUF) and the Polish zloty (PLN) are quite strongly correlated

with each other (see Figure 1) and also with the Czech crown (CZK). Naturally, the dynamic conditional correlation models should provide deeper (and especially dynamic) insight.

$\hat{\mathbf{R}}$	CZK	GBP	HUF	PLN	RON	SEK
CZK	1.000	-0.056	0.393	0.416	0.188	0.169
GBP	-0.056	1.000	-0.044	-0.027	0.014	0.074
HUF	0.393	-0.044	1.000	0.664	0.446	0.321
PLN	0.416	-0.027	0.664	1.000	0.412	0.361
RON	0.188	0.014	0.446	0.412	1.000	0.217
SEK	0.169	0.074	0.321	0.361	0.217	1.000

Table 5: The CCC estimator of the conditional covariance matrix $\mathbf{R}$.

4 Conclusion

In the presented case of the portfolio of six EU currencies, it was observed that the DCC and cDCC estimators are quite similar. Furthermore, the CCC estimator could be basically viewed as (concise) information about an average level of correlation links. With regard to the previously mentioned comparison of the models, it is natural to prefer the cDCC model with its consistent correlation targeting to the DCC (or CCC) one. On the other hand, there is a certain difference between two estimation methods, namely MLE and composite MLE. These findings will be a subject of further research.

From the economic point of view, several interesting observations were performed. For example, the British pound seems to have only weak correlations with the other currencies in the portfolio. Moreover, the currencies of Visegrad countries show stronger correlation links, especially in the case of the Hungarian forint and the Polish zloty.

Acknowledgements

This work was supported by Czech Science Foundation (the grant P402/12/G097). The author also wishes to express his thanks to Prof. RNDr. Tomáš Cipra, DrSc. for several helpful comments.

References

[1] Aielli, G. P.: *Dynamic Conditional Correlation: On Properties and Estimation.* Working paper available at SSRN: http://ssrn.com/abstract=1507743, 2011.

[2] Bauwens, L., Laurent, S. and Rombouts, J. V. K: Multivariate GARCH Models: A Survey, *Journal of Applied Econometrics* **21** (2006), 79-109.

[3] Bollerslev, T.: Modeling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model, *Review of Economics and Statistics* **72** (2000), 498-505.

[4] Engle, R.: *Anticipating Correlations: A New Paradigm for Risk Management.* Princeton University Press, Princeton, 2009.

[5] Engle, R.: Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroscedasticity Models, *Journal of Business & Economic Statistics* **20** (2002), 339-350.

[6] Engle, R. and Sheppard, K.: *Theoretical and Empirical properties of Dynamic Conditional Correlation Multivariate GARCH.* NBER Working Paper No. 8554 available at http://www.nber.org/papers/w8554.pdf, 2001.

The Credit Market Model with Three Parameters

Tomáš Heryán[1]

Abstract: We present a new theoretical model of the credit market in current paper. Our analyzing of earlier studies in this research area has proved that all the authors worked with three parameters of the credit market. As the first, interest rate on loans that means risk payment of the loan contract. Second, the creditworthiness or the bonity that means the quality of the credit client. And third, the credit collateral which is sometimes given to secure the loan contract. The aim of this study is to construct new credit market model with all these three parameters. All functions in this model are mathematically described. In addition, we modify our basic model due to asymmetric information problematic. Even if the mathematical confirmation of the potential performance of the model is included in this paper, we have highlighted some theoretical barriers why the credit market model could not work in the real economy. Despite the fact we argue that the type of our theoretical model can be *ceteris paribus* very close to the practical functioning of the credit market in the real economy. The present study is the basis for the further research work in the credit market area in future.

Keywords: Credit market, Mathematical functions, Theoretical model, Information asymmetry.

JEL Classification: B41, D53, G10
AMS Classification: 26B06

1 Introduction

Theoretical models of the credit market were described during second half of the 20th century. Authors went through basic models in e.g. Jaffee [5] to models affected by information asymmetry as in e.g. Stiglitz, Weiss [6]. Several studies examine credit market applying some from three variables. As the first, interest rate on loans that means risk payment of the loan contract. Second, the creditworthiness or the bonity that means the quality of the credit client. And third, the credit collateral which is sometimes given to secure the loan contract. We have not founded any study that present theoretical model of the credit market with all of these variables the way used here. This fact motivates the present study. The present study focused on the credit market theoretical modeling.

The aim of the paper is construct the credit market model with three parameters. Its functions are mathematically described to improve understanding of the model. Contribution to current literature should be in applying all of three variables in another way in the model, definitely.

The paper is organized as follows: In the next section we shortly analyze several studies. In section three we define function assumptions of the credit market model with three parameters. Section four is our theoretical model with mathematically descriptions of its function. In addition we construct advanced theoretical model in next section where we apply information asymmetry. Section six describes model's weaknesses in the real economy context. Section seven it concludes.

2 Literature review

Jaffee [5] described credit market as a perfect competition market where exist non-profit/loss point. This point is a tangent of credit supply and credit demand. Due to credit rationing when credit demand leads credit supply banks achieve profits in equilibrium. It is match of demand and supply. Interest rate on loans is bigger than in non-profit point and banks supply less volume of loans in that. Cosci [3] argue that credit market is close to an oligopoly market than a prefect competition market. She also argues that due to this fact information asymmetry could be beneficial for bank in some cases. Credit market is not monopoly market even if bank is in the role of price maker, definitely in [3] and [5]. Clemenz [2] argue that bank sell loans on higher price than it is optimum interest rate, too. Price maker's role is unusual for an oligopoly market, but that type is the closest to credit market.

[1] Internal Ph.D. student, Department of Finance at Silesian University, School of Business Administration in Karvina, Univerzitní náměstí 1934/3, Karviná, 733 40, Czech Republic, E-mail: heryan@opf.slu.cz.

Stiglitz and Weiss [7] examine more optimal interest rate on loans. They described that as functional relationship between banks income and loan's interest rate. Till the optimal level of interest rate it is increase function. Higher interest rate means higher income of banks. But interest rate means also risk of credit contract. Therefore higher interest rate than optimal interest rate is connected with some losses and function change to decrease. Higher interest rate than optimal means less bank income then. Stiglitz and Weiss [6] in their younger study apply optimal interest rate for whole credit market. Macro view on this problematic is their contribution to the literature.

Townsend [8] examines insurance market but his paper focused on information asymmetry and avoidable moral hazard. He looks on the economy as the system with two agents and one random variable. He characterized optimal contracts applying mathematical descriptions under function assumptions. Bester [1] show that no credit rationing will occur in the credit market equilibrium if banks compete by choosing collateral requirements and interest rate on loans to screen investors' riskiness. In his study he went through credit market with perfect information to the imperfect markets with information asymmetry, concretely adverse selection. Williamson [9] applies asymmetry informed lenders as well as borrowers and established a link between the credit rationing, the equilibrium and the financial intermediation.

In their studies Clemenz [2] and Williamson [9] independently examine credit market with information asymmetry and monitoring costs' problematic. Frexias and Rochet [4] also apply information asymmetry. In their study they examine credit market and also analyze recent studies [1], [7], [8], [9]. Their contribution to the current literature, except completing recent literature, is in applying more credit collateral's problematic.

3 Function assumptions of the model

All theoretical models should be constructed on some assumptions to the function of that, *ceteris paribus*. Therefore we make also function assumptions of our basic model which we will change only a little. These assumptions are:

- Commercial banks do not exceed optimal level of interest rate on loans i_{max} through they maximize their income R_{max}, Stiglitz and Weiss [7].
- Banks do not exceed also the lowest level of interest rate on loans i_{min} to achieve minimum income R_{min}. In our case it is not a perfect competition market as in Jaffee [5].
- Loan's interest rate is determined by assessing the bank client's creditworthiness or bonity b and loan's collateral C, Freixas and Rochet [4].
- The credit market is lending through the match or intersection of the credit supply S_L and credit demand D_L. The supply is exogenous variable (given by banking sector) and the demand is endogenous variable (dependent on banking sector, especially on interest rates on loans).
- Only perfect information exists on the credit market in the basic theoretical model. Therefore interest rate on loans is a good measure of the credit risk from lending process.

4 Basic theoretical model of credit market

First, the basic theoretical model of the credit market with three parameters is mathematically described. Second, we construct Cartesian graph to illustrate mathematic functions and interconnections between credit supply S_L and credit demand D_L.

Credit supply S_L in the basic model is mathematically described by next equations:

$$L = S_L(i) \tag{1}$$
$$i = f(R) \tag{2}$$
$$L_r = g(R) \tag{3}$$

, where L is increasing function of interest rate on loans i and $S_L(i)$ means loans volume in (1). Increasing function of bank's profit $f(R)$ means level of interest rate on loans i in (2). On the other hand, increasing function of bank's profit $g(R)$ means also level of the credit risk L_r in (3). That means relationships through we can construct credit supply S_L in figure 1 then: $i \rightarrow L,\ R \rightarrow i,\ R \rightarrow L_r$.

Credit demand D_L in the basic model is mathematically described by next equations:

$$L = D_L(i) \tag{4}$$
$$i = f(R) \tag{5}$$
$$R = h(C\&b) \tag{6}$$
$$C\&b = k(C, b) \tag{7}$$

, where inverse functional relationship between interest rate on loans i and the loans' volume D_L, means decreasing function L in (4). Inverse increasing function f^1 to function f of interest rate on loans i in (5). Decreasing function h of the collateral and the creditworthiness $C\&b$, means safe income from lending R in (6). Function k means just coherence between credit collateral and credit bonity. On these relationships we can construct credit demand D_L in figure 1 then: $i \rightarrow L,\ R \rightarrow i,\ C\&b \rightarrow R$.

The basic credit market model with three parameters is mathematically described by next equations then:

$$L_S = S_L\left(f\left(g^{-1}(L_r)\right)\right) \tag{8}$$

$$L_D = D_L\left(f\left(h(C\&b)\right)\right) \tag{9}$$

$$S_L\left(f\left(g^{-1}(L_r)\right)\right) = D_L\left(f\left(h(C\&b)\right)\right) \tag{10}$$

$$L_r = g\left(f^{-1}\left(S_L^{-1}\left(D_L\left(f\left(h(C\&B)\right)\right)\right)\right)\right) \tag{11}$$

, where we can make through relationships (1) – (3) loan's quantity offered in the market L_S in (8). We can make loan's quantity inquired in the market L_D in (9) from relationships (4) – (6). On an equal basis and match of credit supply S_L and credit demand D_L we make mathematical output description in (10). From the equal basis relationship (10) we make also relationship between credit risk L_r and credit collateral, bonity $C\&b$ in (11).

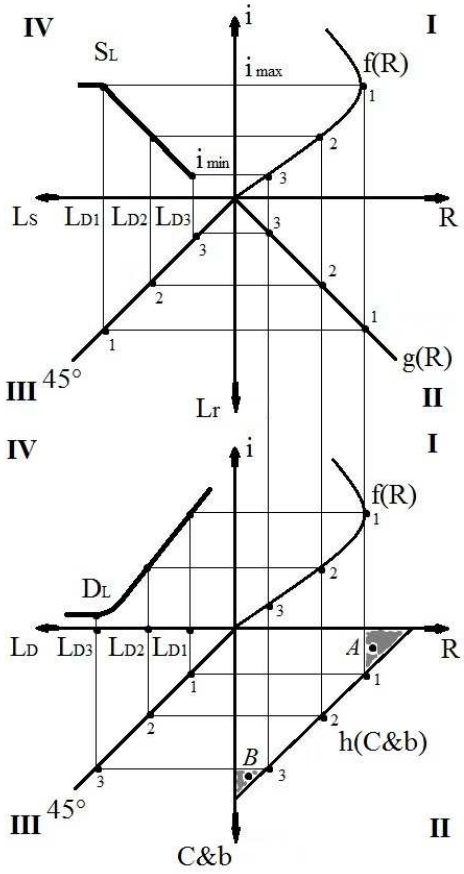


Figure 1 Basic theoretical model of credit market with three variables

On figure 1 we can see Cartesian graph with the line 45° [3], defining the credit market's supply and demand by three selected factors. In the first quadrant of the graph with the credit supply, we can see the function f of the proceeds of the loan transactions, described by [7]. The maximum rate we consider optimal interest rate on loans i_{max}. On a higher level of interest rate f changes into inverse function of the nature (breaking point). The second quadrant expresses the quality of the credit transaction L_r by function g. It is characterized by three parameters. This reflects the creditworthiness of the bank's client here, the parameters of the product, but also the quality of credit collateral. The second quadrant corresponds to the credit supply with a view of reality banks. More liquid loans are more profitable because the bank can continue with the funds to operate. For loans with longer maturities the interest rate is also at a lower level. As liquidity in the credit business we understand its potential to change back to the funds that the debtor bank borrowed. By increasing function g of the credit quality of trade is

enter several factors. The right half of the graph represents the investment magically triangle from the credit supply side. First risk equals interest rate, second the yield on the loan and credit liquidity contract then. After the projection through the third quadrant, the line of 45° we are able to model the supply curve in the credit market in the last fourth quadrant.

The first quadrant of the second half Cartesian´s graph, demand of the credit market is identical to the first half, the supply side with the same function f. The second quadrant, the function h is the theoretical basis described by [4]. The less creditworthy clients with less collateral are providing a higher interest rate than more creditworthy or more secure clients. Therefore it constitutes a higher risk for the bank. The right half of the graph it is a magical investment triangle on the demand for loans again. Risk equals interest rate on loans, the yield on loans and liquidity is given by client bonity b or the quality of collateral C. After the projection through the third quadrant, with the line of 45° again we are able to model the credit demand curve in the last fourth quadrant.

In the case of the credit offers exists an increasing supply law. At a higher level of interest rate on loans banks offer more loans. At the same time the borrower at a lower quality (lower credit quality or lower quality of collateral) interest rate increases. On demand side exist on the other hand a decreasing demand law. Where, a higher level of interest rate on loans decreases inquiries for loans. At the same time, however, from the perspective of banks, clients are able to buy only adequate quality. Therefore, in the second quadrant of the graph indicated the inquiry points A, B, located in different gray areas. Clients A are looking a little reputable banks and bonity b or are unable to provide security C in sufficient amount and quality and due to they cannot reach the credit. The bank would also such clients have not been able to determine the appropriate interest rate risk of the contract in which reach revenues. Clients B are *vice-versa* creditworthy, it would still deserve a lower interest rate than i_{min} but the bank would not achieve the minimum expected returns and would not meet one of function assumptions of the model. These clients are therefore forced to buy a minimum interest rate set by the bank or the supply and demand does not occur.

On figure 1 there are three types of credit contracts. Borrower 1 is more risky for the bank. He has the interest rate on the loan at its maximum, optimal level i_{max}. We can also see that such debtor 1 is granted a loan at the lowest level L_{D1}. With decreasing the risk of further decreasing the interest rate on loans established by the bank, and an increased amount of credit provided by the client 2. He reaches a higher amount of money L_{D2} at a lower price (or in the bank's portfolio there are more borrower clients 2 than type 1). Borrower 3 is very specific, because in terms of its credit risk they are offered at the lowest interest rate level i_{min}. Such a borrower can reach the largest amount of funds L_{D3} at the lowest price then (or such debtors are in bank's portfolio more than types of borrowers 1, 2). However, in terms of credit risk the bank paid above the facts relating to its portfolio of credit clients. The bank can maximize its revenues by determining the optimal interest rate, as described by [7] only in setting a rate for each loan contract separately.

Another important feature of basic theoretical model of the credit market with three parameters is a break in the case of supply curve [5], and demand curve. Maximum interest rate i_{max} is determined based on maximizing income rules of the bank from function assumptions of the model. It does not mean that in reaching this maximum limit banks will not be further lending. However, offer becomes perfectly elastic up to a maximum amount offered. Also set the lower limit of interest rates i_{min} is based on minimum expected returns of banks from function assumptions. That does not mean that credit clients will stop lending at a lower rate demands, too. Quantity is rising at a lower cost of funds and a lot more demand curve becomes perfectly elastic. The minimum investment rate of return is determined by the market because after all just an interest rate on loans still exists.

5 Advanced theoretical model – information asymmetry

The basic model of credit market with three variables should be modified due to information asymmetry problematic. Recent studies deal with that Bester [1], Cosci [3], Frexias and Rochet [4], Stigltz and Weiss [6], [7]. Due to adverse selection or moral hazard all our selected variables are not correctly determined in real economy. We break one defined function assumption of the model and apply information asymmetry in model advanced.

The advanced theoretical model with information asymmetry is mathematically described by these equations:

$$L_S = L_D + \delta \tag{12}$$

$$S_L\left(f\left(g^{-1}(L_{r\delta})\right)\right) = D_L\left(f\left(h(C\&b)\right)\right) + \delta \tag{13}$$

$$L_{r\delta} = g\left(f^{-1}\left(S_L^{-1}\left(D_L\left(f\left(h(C\&b)\right)\right) + \delta\right)\right)\right) \tag{14}$$

$$L_{r\delta} = L_r + \frac{d}{d(C\&b)}\, g\left(f^{-1}\left(S_L^{-1}\left(D_L\left(f\left(h(C\&b)\right)\right)\right)\right)\right).\delta + \varepsilon \tag{15}$$

, where δ means the risk of information asymmetry in credit market in (12). From (10) we modify (13), and from (11) we modify (14) then. Finally we apply Taylor expansion to make (15).

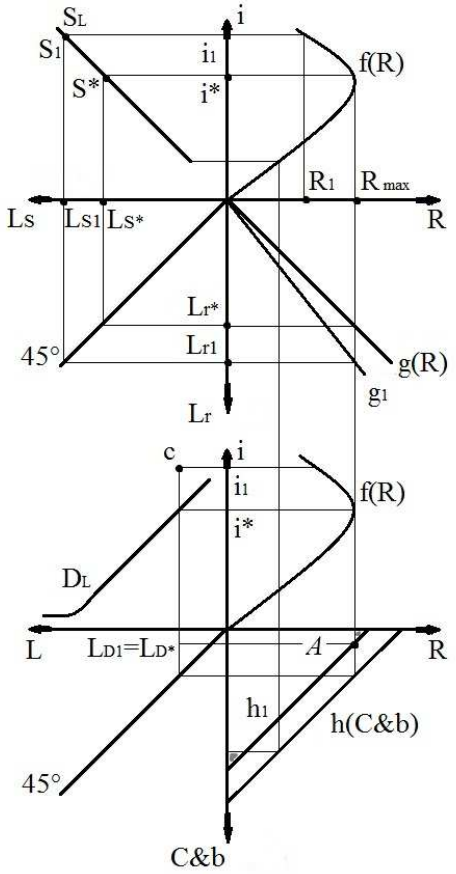


Figure 2 Advanced theoretical model of credit market with three variables

If there is ex ante adverse selection, the bank asses wrong the credit quality of the business for example. Of course, this could be affected by only one of the following factors. Furthermore, we describe all the factors that could cause adverse selection. Already in the second quadrant of the graph is an invitation to cross to the quality of trade credit, bank given set of function g is different in fact. This is a less trade quality, and therefore the derived function is given as more inelastic g_1. This leads to a situation in which the bank achieve the expected return R_{max}, it would have real liquidity trade to increase from the original L_r* level to L_{r1}. Another possible cause of adverse selection is on the demand side in less than a bank specified quality of the credit client. In the second quadrant of the graph so the inquiry is to shift the entire curve of function h. The creditworthiness or bonity b and the quality of collateral C is at a lower level of the derived function h_1. It diminishes both the gray zones and in the previous case. We see that the client A, located in the previous case in poor gray zone becomes offered by effect of adverse selection. But the bank does not know about adverse selection because of a lower quality credit contract valued as in the previous case. Line 45° showing a steady relationship is exceeded in the third quadrant and in the last quadrant so we get into a point c lying above the level of demand [3]. We see that such a contract should be rated higher than the optimal interest rate on loans i_{max}. So the yield curve of the function f we obtain the loss making parts and the function f becomes inverse character. In the case of credit supply, although we see that the third quadrant 45° line intersects the steady point more distant from the center. However, asymmetric information causes that the bank actually offer more than the optimal amount of loans at lower interest rates. It means inadequate risk for the credit contracts. Due to information asymmetry supply curve is not in the optimum point perfectly elastic. When projected supply real interest rates, higher than optimal, we obtain the *R-axis* yields a loss caused ex ante by adverse selection, where $R_1 < R_{max}$.

In another situation when moral hazard exist on the credit market, the credit relationship becomes an ex post lower quality. Whether, it is a deterioration of the credit business of the function f or lower client quality of the function h. It will have exactly the same impact on the credit market as in the case of adverse selection [6]. For this reason we have poor client and leave the inferior function h_1. In fact, A should lie outside the curve but still outside the gray zone. In terms of moral hazard interpretation, by contrast, only we can say that when it got into the relationship loan balance states except the bank's monitoring costs are the same. Provided a greater than

optimal quantity offered credit, the credit risk do not equals interest rate. So again we get to the point c, which is part of a larger than optimal area and lie above the credit demand in the credit market.

6 Weaknesses of the model in the real economy

We make also some weaknesses of the credit market model with three parameters and its function in the real economy. These are:

- Each bank has its own clients with different risk L_r and cannot achieve the maximum returns R_{max} from credits.
- Due to bank's competition some bank cannot determine level of optimal interest rate i_{max} on loans independently on the other banks and their level of interest rate on loans.
- Information asymmetry δ is very difficult to measure or even predict the both, adverse selection and moral hazard.
- When information asymmetry impact the credit market, level of the credit risk L_r change immediately and the optimal level of interest rate on loans i_{max} change, too.

7 Conclusion

The aim of the paper was to construct the credit market model with three parameters. Its functions are mathematically described to improve understanding of the model. We construct also advanced model where we applied information asymmetry problematic.

Discussion on theoretical results really proved that this type of the credit market model could be useable for analyzing the credit markets in some countries. It is important differ between types of financial systems then. This is just primary research in this area. It should be complete with the numerical example that could be tested applying panel regression in the future.

We proved that all three parameters, interest rate on loans, creditworthiness and the credit collateral, have some relationships and impact to each other. We cannot skip some of them for detail analyzing of the credit market. Problematic of information asymmetry should not be underestimated, too.

Acknowledgements

Research behind this paper was supported by the Student Grant Competition of Silesian University within the project SGS 25/2010 'Financial integration in the EU and its effect on corporate sector'.

References

[1] Bester, H.: Screening vs rationing in credit markets with imperfect information. *American Economic Review.* **75** (1985). 850–855.
[2] Clemenz, G.: *Credit Markets with Asymmetric Information.* Springer-Verlag, Berlin, 1986.
[3] Cosci, S.: *Credit Rationing and Asymmetric Information.* Dartmouth Publishing Company, Hantz, 1993.
[4] Frexias, X., Rochet, J.C., *Microeconomics of Banking.* 2nd Edition, The MIT Press, Cambridge, 2008.
[5] Jaffee, D.M. *Credit Rationing and the Commercial Loan Market: An Econometric Study of the Structure of the Commercial Loan Market.* John Wiley & Sons Inc., London, 1971.
[6] Stiglitz, J.E., Weiss, A. Asymmetric Information in Credit Markets and Its Implications for Macro-Economics. *Oxford Economic Papers.* **44** (1992), 694-724.
[7] Stiglitz, J.E., Weiss, A. Credit Rationing in Markets with Imperfect Information. *American Economic Review.* **71** (1981), 393-410.
[8] Townsend, R.M. Optimal Contracts and Competitive Markets with Costly State Verification. *Journal of Economic Theory.* **21** (1979), 265-293.
[9] Williamson, S.D. Costly Monitoring, Financial Intermediation and Equilibrium Credit Rationing. *Journal of Monetary Economy.* **18** (1986), 159-179.

An interval linear programming contractor

Milan Hladík[1]

Abstract. We consider linear programming with interval data. One of the most challenging problems in this topic is to determine or tight approximate the set of all optimal solutions subject to all perturbations within the given intervals. We propose an iterative method that finds an enclosure of the set of optimal solutions. The method is based on a linear approximation and sequential refinement. It runs in polynomial time, so, naturally, convergence to the ideal set cannot be ensured. We apply the method in a simple portfolio selection problem with uncertain data.

Keywords: interval linear programming, interval analysis

JEL classification: C44
AMS classification: 90C31

1 Introduction

Since linear programming is widely used in modelling real-life problems, it must take into account inaccuracies and data measurement errors, which are common in most of the problems. There are plenty of various approaches to handle uncertainty in linear systems, see e.g. [4]. In this paper, we deal with an interval linear programming, in which we assume that there are *a priori* known some intervals in which inexact quantities may perturb. Linear programming with interval data has been studied for forty years; see a survey paper [10]. The problems discussed are the optimal value range [4], [7], [12], basis stability, and duality [5], among others. Interval linear programming was applied in portfolio selection problems [6], environmental management [15] interval matrix games [16], and can also serve in fuzzy linear regression as an alternative to traditional approaches [11, 3].

In this paper, we focus on the optimal solution set. The problem of calculating the set of all possible solutions over all data perturbations is considered to be very difficult. It becomes tractable in the special case of the so called basis stability [8], [14], [17], meaning that there is a basis that is optimal under any admissible perturbation. However, basis stability is not so easy to verify; indeed, it is an NP-hard problem [10]. Moreover, since many practical problems suffer from degeneracy, one cannot expect that basis stability holds true in general.

Thus, the research was driven to calculate an enclosure (interval superset) of the optimal solution set. Such enclosures can be computed e.g. by using interval arithmetic [2], [13], instead of the real one, but the results are usually very overestimated.

Our aim is to propose an efficient algorithm for computing an enclosure of the optimal solution set. We present an iterative algorithm that starts with an initial enclosure and sequentially makes it tighter. Naturally, it doesn't converge to the optimal enclosure in general, but (in view of the performed examples) it gives bounds that are sufficiently accurate for many purposes.

Let us introduce some notations. An interval matrix is a family of matrices

$$\boldsymbol{A} := \{A \in \mathbb{R}^{m\times n} \mid \underline{A} \leq A \leq \overline{A}\},$$

where $\underline{A}, \overline{A} \in \mathbb{R}^{m\times n}$ are given. The midpoint and the radius of an interval matrix $\boldsymbol{A}$ is defined respectively as

$$A^c := \frac{1}{2}(\underline{A} + \overline{A}), \quad A^\Delta := \frac{1}{2}(\overline{A} - \underline{A}),$$

[1] Charles University, Faculty of Mathematics and Physics, Department of Applied Mathematics, Malostranské nám. 25, 118 00, Prague, Czech Republic, and
University of Economics, Faculty of Informatics and Statistics, nám. W. Churchilla 4, 13067, Prague, Czech Republic, e-mail: milan.hladik@matfyz.cz

and similarly for interval vectors. The set of all interval $m \times n$ matrices is denoted by $\mathbb{IR}^{m\times n}$. A solution to the interval system $\boldsymbol{A}x = \boldsymbol{b}$, $x \geq 0$ means any solution to any scenario $Ax = b$, $x \geq 0$ with $A \in \boldsymbol{A}$, $b \in \boldsymbol{b}$. Similarly for interval systems of inequalities. The diagonal matrix with entries $z_1, \ldots, z_n$ is denoted by $\mathrm{diag}(z)$.

2 Interval linear programming

Consider a linear programming problem

$$\min c^T x \;\text{ subject to }\; Ax = b,\; x \geq 0, \tag{1}$$

and its dual

$$\max b^T y \;\text{ subject to }\; A^T y \leq c.$$

Suppose that (possibly all) input values are inexact, and we have lower and upper limits for the ranges of variations. Thus, we are given $\boldsymbol{A} \in \mathbb{IR}^{m\times n}$, $\boldsymbol{b} \in \mathbb{IR}^m$, $\boldsymbol{c} \in \mathbb{IR}^n$, and we focus on the family of linear programs (1) with $A \in \boldsymbol{A}$, $b \in \boldsymbol{b}$, $c \in \boldsymbol{c}$. Denote by $\mathcal{S}(A, b, c)$ the set of optimal solutions to (1). By the set of optimal solutions to an interval linear program we understand the set

$$\mathcal{S} := \bigcup_{A\in\boldsymbol{A},\, b\in\boldsymbol{b},\, c\in\boldsymbol{c}} \mathcal{S}(A, b, c).$$

The set $\mathcal{S}$ needn't be a polyhedron, and it is hard to determine it in general. We want to give a cheap calculation for an enclosure to $\mathcal{S}$.

By using duality theory, we have that $x \in \mathcal{S}$ if and only if there is some $y \in \mathbb{R}^m$, $A \in \boldsymbol{A}$, $b \in \boldsymbol{b}$, and $c \in \boldsymbol{c}$ such that

$$Ax = b,\; x \geq 0,\; A^T y \leq c,\; c^T x = b^T y. \tag{2}$$

Relaxing the dependency we obtain a superset described by

$$\boldsymbol{A}x = \boldsymbol{b},\; x \geq 0,\; \boldsymbol{A}^T y \leq \boldsymbol{c},\; \boldsymbol{c}^T x = \boldsymbol{b}^T y. \tag{3}$$

Notice that the solution sets are not equal, since in the latter the intervals vary independently within $\boldsymbol{A}$ and $\boldsymbol{A}^T$, while in the former they are related. By [10], the solution set to (3) is described as

$$\underline{A}x \leq \overline{b},\; -\overline{A}x \leq -\underline{b},\; x \geq 0,\; A_c^T y - A_\Delta^T |y| \leq \overline{c},\; |c_c^T x - b_c^T y| \leq c_\Delta^T x + b_\Delta^T |y|. \tag{4}$$

It represents a non-convex polyhedron, which becomes convex when we restrict the signs of y_i, $i = 1, \ldots, m$. A simple method introduced in [10] was based on decomposing (4) into 2^m sub-problems according to the signs of y_i, $i = 1, \ldots, m$. Each sub-problem has a linear description and is solved by ordinary linear programming. However, the method is very time consuming as m grows.

In the following, we employ a kind of linearization of (4), which is based on a result by Beaumont [1]. It needs an initial enclosure, which is used for refinement.

Theorem 1 (Beaumont, 1998)**.** *For every $y \in \boldsymbol{y} \subset \mathbb{R}$ with $\underline{y} < \overline{y}$ one has*

$$|y| \leq \alpha y + \beta, \tag{5}$$

where

$$\alpha = \frac{|\overline{y}| - |\underline{y}|}{\overline{y} - \underline{y}} \quad \textit{and} \quad \beta = \frac{\overline{y}|\underline{y}| - \underline{y}|\overline{y}|}{\overline{y} - \underline{y}}.$$

Moreover, if $\underline{y} \geq 0$ or $\overline{y} \leq 0$ then (5) holds as equation.

Let $\boldsymbol{x} \in \mathbb{IR}^n$ and $\boldsymbol{y} \in \mathbb{IR}^m$ be an enclosure to (4). By the Beaumont theorem, we will linearize the absolute value $|y|$ as follows. Define vectors $\alpha, \beta \in \mathbb{R}^m$ componentwise as

$$\alpha_i := \begin{cases} \frac{|\overline{y}_i| - |\underline{y}_i|}{\overline{y}_i - \underline{y}_i} & \text{if } \underline{y}_i < \overline{y}_i, \\ \mathrm{sgn}(\overline{y}_i) & \text{if } \underline{y}_i = \overline{y}_i, \end{cases}$$

$$\beta_i := \begin{cases} \frac{\overline{y}_i|\underline{y}_i| - \underline{y}_i|\overline{y}_i|}{\overline{y}_i - \underline{y}_i} & \text{if } \underline{y}_i < \overline{y}_i, \\ 0 & \text{if } \underline{y}_i = \overline{y}_i. \end{cases}$$

Then the linearization of (4) reads

$$\underline{A}x \le \overline{b},\ -\overline{A}x \le -\underline{b},\ x \ge 0, \tag{6a}$$

$$\big(A_c^T - A_\Delta^T \operatorname{diag}(\alpha)\big)y \le \overline{c} + A_\Delta^T\beta, \tag{6b}$$

$$\underline{c}^T x + \big(-b_c^T - b_\Delta^T \operatorname{diag}(\alpha)\big)y \le b_\Delta^T\beta, \tag{6c}$$

$$-\overline{c}^T x + \big(b_c^T - b_\Delta^T \operatorname{diag}(\alpha)\big)y \le b_\Delta^T\beta. \tag{6d}$$

Now, we compute the interval hull to (6). If it is smaller than the initial enclosure $\boldsymbol{x}, \boldsymbol{y}$ then we can iterate the process to obtain more tight enclosure. It is a basic idea of our method described in Algorithm 1.

The initial enclosure $\boldsymbol{x}^0, \boldsymbol{y}^0$ from step 1 is taken as

$$\boldsymbol{x}^0 := ([0,K],\dots,[0,K])^T, \quad \boldsymbol{y}^0 := ([-K,K],\dots,[-K,K])^T,$$

where $K \gg 0$ is large enough. The stopping criterion used in step 7 is

$$\frac{\sum_{j=1}^n (x_\Delta^i)_j + \sum_{j=1}^m (y_\Delta^i)_j}{\sum_{j=1}^n (x_\Delta^{i-1})_j + \sum_{j=1}^m (y_\Delta^{i-1})_j} \ge 0.99.$$

Algorithm 1 (Optimal solution set contractor)

1: Compute an initial interval enclosure $\boldsymbol{x}^0, \boldsymbol{y}^0$ of (4);
2: $i := 0$;
3: **repeat**
4: compute α and β by using $\boldsymbol{y}^i$;
5: $i := i + 1$;
6: compute the interval hull $\boldsymbol{x}^i, \boldsymbol{y}^i$ of (6);
7: **until** improvement is nonsignificant;
8: **return** $\boldsymbol{x}^i$;

Remark 1. There is a natural problem how to choose a sufficiently large K for the initial enclosure. Even though the algorithm works well for very conservative K (cf. Example 1), in some cases, we can validate that K is large enough.

Suppose that $\boldsymbol{x}^0, \boldsymbol{y}^0$ contain at least one optimal solution for some scenario (which is easy to satisfy). Next, suppose that (2) is solvable for each $A \in \boldsymbol{A}$, $b \in \boldsymbol{b}$, and $c \in \boldsymbol{c}$; this issue is discussed e.g. in [10]. Take the initial enclosure $\boldsymbol{x}^0, \boldsymbol{y}^0$ as

$$\boldsymbol{x}^0 := ([-1,K],\dots,[-1,K])^T, \quad \boldsymbol{y}^0 := ([-K,K],\dots,[-K,K])^T,$$

If the interval vectors $\boldsymbol{x}^1, \boldsymbol{y}^1$ from the next iteration are strictly inside the initial ones, that is, if $\underline{x}^0 < \underline{x}^1 \le \overline{x}^1 < \overline{x}^0$ and $\underline{y}^0 < \underline{y}^1 \le \overline{y}^1 < \overline{y}^0$, then $\boldsymbol{x}^1, \boldsymbol{y}^1$ comprise all optimal solutions (this is why we put the lower limits of $\boldsymbol{x}^0$ as -1 instead of 0). This is easily seen from the continuity reasons since, under our assumption, the optimal solution set is connected.

Example 1. Consider an interval linear program

$$\begin{aligned} \min -[15,16]x_1 - [17,18]x_2 \ \text{ subject to} \\ x_1 &\le [10,11], \\ -x_1 + [5,6]x_2 &\le [25,26], \\ [6,6.5]x_1 + [3,4.5]x_2 &\le [81,82], \\ -x_1 &\le -1, \\ x_1 - [10,12]x_2 &\le -[1,2]. \end{aligned}$$

Even though it is not in the standard form (1), the associated primal–dual pair is the same as for (1). We take the initial enclosure

$$\begin{aligned} \boldsymbol{x}^0 &= 1000 \cdot ([-1,1],\,[-1,1])^T, \\ \boldsymbol{y}^0 &= 1000 \cdot ([0,1],\,[0,1],\,[0,1],\,[0,1],\,[0,1])^T. \end{aligned}$$

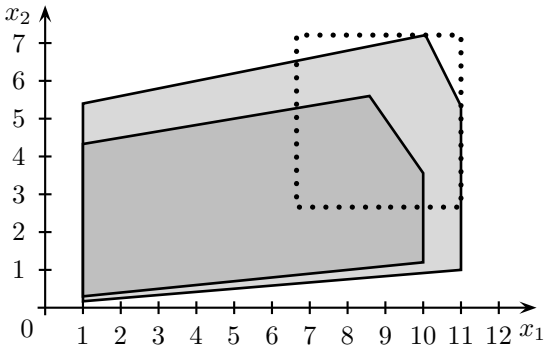


Figure 1: Intersection of all feasible sets in dark gray; union in light gray; the resulting enclosure of optimal solutions represented by the dotted rectangle.

The iterations of the procedure go as follows

$$\begin{aligned}
\boldsymbol{x}^1 &= ([1,11],\,[-568,916])^T,\\
\boldsymbol{y}^1 &= ([0,1000],\,[0,936],\,[0,358],\,[0,1000],\,[0,572])^T,\\
\boldsymbol{x}^2 &= ([1,11],\,[-17.2,72])^T,\\
\boldsymbol{y}^2 &= ([0,190],\,[0,58.5],\,[0,24.3],\,[0,176],\,[0,34.6])^T,\\
\boldsymbol{x}^3 &= ([3.78,11],\,[1.91,9.80])^T,\\
\boldsymbol{y}^3 &= ([0,30.6],\,[0,6.98],\,[4.71],\,[0,17.1],\,[0,3.09])^T,\\
\boldsymbol{x}^4 &= ([6.65,11],\,[2.66,7.21])^T,\\
\boldsymbol{y}^4 &= ([0,22.5],\,[0.08,4.33],\,[0,3.67],\,[0,8.81],\,[0,1.47])^T.
\end{aligned}$$

The fifth iteration does no improvement, so we return the enclosure $\boldsymbol{x}^4$. Notice that the same result is obtained by the exponential decomposition procedure discussed above, but with a much more computational effort. An illustration is given in Figure 1. We see that the upper bounds are very tight, but the lower bounds are quite conservative.

Example 2. Let us apply our approach in a simple portfolio selection model. Consider we have J possible investments for T time periods and r_{tj}, $t = 1,\dots,T$, $j = 1,\dots,J$, stands for return on investment j in time period t. Estimated reward on investment j using historical means is defined as

$$W_j := \frac{1}{T}\sum_{t=1}^{T} r_{tj}, \quad j = 1,\dots,J,$$

or in matrix form $W = \frac{1}{T}R^T e$, where e is the vector of ones (with convenient dimension). In order to get a linear programming formulation of the problem we measure risk of investment j by sum of absolute values instead of the sum of the squares:

$$\frac{1}{T}\sum_{t=1}^{T} |r_{tj} - W_j|.$$

Let μ be a risk aversion parameter (upper bound for risk) given by a user, and the variables x_j, $j = 1,\dots,J$, denotes a fraction of portfolio to invest in j. Then the maximal allowed risk is expressed by the constraint

$$\frac{1}{T}\sum_{t=1}^{T} \Big|\sum_{j=1}^{J}(r_{tj} - W_j)x_j\Big| \le \mu,$$

or, by converting to a linear inequality system

$$-y_j \le \sum_{j=1}^{J}(r_{tj} - W_j)x_j \le y_t,\ \forall t = 1,\dots,T, \quad \frac{1}{T}\sum_{t=1}^{T} y_t \le \mu.$$

In a compact matrix form, it reads

$$-y \le \Big(I - \frac{1}{T}E\Big)Rx \le y, \quad \frac{1}{T}e^T y \le \mu,$$

where E is the matrix of ones and I the identity matrix. In order to obtain a robust formulation of the portfolio selection problem, we state the linear programming problem as

$$\begin{aligned} \max \sum_{j=1}^{J} W_j x_j = \frac{1}{T}e^T Rx \\ \text{subject to } -y \le \Big(I - \frac{1}{T}E\Big)z \le y, \quad Rx = z, \\ e^T x = 1, \quad \frac{1}{T}e^T y \le \mu, \quad x, y \ge 0. \end{aligned}$$

For concreteness, consider a portfolio selection problem with $J = 4$ investments and $T = 5$ time periods. The risk aversion parameter is set as $\mu := 2$. The returns are displayed below:

time period t	reward on investment			
	1	2	3	4
1	10	20	9	11
2	12	25	11	14
3	9	17	12	12
4	11	21	11	14
5	11	19	13	16

The optimal solution is $x^* = (0,\, 0.9643,\, 0.0357,\, 0)^T$ and the corresponding optimal return is ≈ 20.08.

Suppose that the returns are not known precisely. We extend the values of r_{tj} to intervals $[0.99r_{tj}, 1.01r_{tj}]$, that is, the returns may vary independently and simultaneously within 1% tolerance. We calculate the following enclosure of the optimal solutions

$$\boldsymbol{x}^{(1)} = ([0, 0.1699],\, [0.7621, 1],\, [0, 0.181],\, [0, 0.2379])^T.$$

Even though the result is very conservative, we can conclude some interesting properties. Regardless the setting of values from intervals, we can be sure that at least 75% of the optimal portfolio is directed to the second investment. The remaining investments constitute at most 17%, 18% and 24% of the portfolio, respectively.

Now, let study 5% perturbations of all returns except for the second investment. Thus, we replace the values r_{jt} by intervals $[0.95r_{tj}, 1.05r_{tj}]$, $j = 1, 3, 4$, $t = 1, \ldots, T$. The calculated enclosure for the optimal solutions is very tight:

$$\boldsymbol{x}^{(2)} = ([0, 0.0495],\, [0.9276, 0.9712],\, [0, 0.0531],\, [0, 0.0724])^T.$$

Notice that this problem setting is not B-stable [9], that is, there is no basis being optimal for each interval realization. For instance, putting the fourth investment returns to the upper limit and the others to the lower limit, we get a solution $x^* = (0,\, 0.9492,\, 0,\, 0.0508)^T$. In this setting, it is optimal to invest in the fourth subject instead of the third one.

3 Conclusion

We proposed a method for contracting an interval enclosure of the optimal solution set. It was based on a linearization and iterative refinement. Even though it doesn't converge to the optimal bounds in general, it gives a sufficiently tight enclosure in short time. Thus it can be used as a method for solving interval linear programming itself, or a first step in more involved algorithms.

The method seems to converge quickly in spite of a very huge initial enclosure. However, to decrease the number of iterations, we will address the future research to finding a more appropriate initial enclosure. We will also carry out more numerical experiments to empirically verify the convergence speed.

Acknowledgements

The author was partially supported by the Czech Science Foundation Grant P403/12/1947.

References

[1] Beaumont, O.: Solving interval linear systems with linear programming techniques. *Linear Algebra and its Applications* **281** (1998), 293–309.

[2] Beeck, H.: Linear programming with inexact data. technical report TUM-ISU-7830, Technical University of Munich, Munich, 1978.

[3] Černý, M.: Computational aspects of regression analysis of interval data. *Proceedings of World Academy of Science, Engineering and Technology* **81** (2011), 421–428.

[4] Fiedler, M., Nedoma, J., Ramík, J., Rohn, J., and Zimmermann, K.: *Linear optimization problems with inexact data*. Springer, New York, 2006.

[5] Gabrel, V. and Murat, C.: Robustness and duality in linear programming. *Journal of the Operational Research Society* **61** (2010), 1288–1296.

[6] Hladík, M.: Tolerances in portfolio selection via interval linear programming. In: *CD-ROM Proceedings 26-th Int. Conf. Mathematical Methods in Economics MME08, Liberec, Czech Republic*, 2008, 185–191.

[7] Hladík, M.: Optimal value range in interval linear programming. *Fuzzy Optimization and Decision Making* **8** (2009), 283–294.

[8] Hladík, M.: How to determine basis stability in interval linear programming. Technical report KAM-DIMATIA Series (2010-973), Department of Applied Mathematics, Charles University, Prague, 2010.

[9] Hladík, M.: Interval linear programming: A survey. Technical report KAM-DIMATIA Series (2010-981), Department of Applied Mathematics, Charles University, Prague, 2010.

[10] Hladík, M.: Interval linear programming: A survey. In: *Linear Programming - New Frontiers in Theory and Applications* (Z. A. Mann, ed.), Nova Science Publishers, New York, 2012, chapter 2, 85–120.

[11] Hladík, M. and Černý, M.: Interval regression by tolerance analysis approach. *Fuzzy Sets and Systems* **193** (2012), 85–107.

[12] Jansson, C.: Rigorous lower and upper bounds in linear programming. *SIAM Journal on Optimization* **14** (2004), 914–935.

[13] Jansson, C.: A self-validating method for solving linear programming problems with interval input data. In: *Scientific computation with automatic result verification* (U. Kulisch, and H. J. Stetter, eds.), Springer, Wien, 1988, 33–45.

[14] Koníčková, J.: Sufficient condition of basis stability of an interval linear programming problem. *ZAMM, Zeitschrift fuer Angewandte Mathematik und Mechanik* **81** (2001), 677–678.

[15] Li, Y. P., Huang, G. H., Guo, P., Yang, Z. F., and Nie, S. L.: A dual-interval vertex analysis method and its application to environmental decision making under uncertainty. *European Journal of Operational Research* **200** (2010), 536–550.

[16] Liu, S.-T. and Kao, C.: Matrix games with interval data. *Computers & Industrial Engineering* **56** (2009), 1697–1700.

[17] Rohn, J.: Stability of the optimal basis of a linear program under uncertainty. *Operations Research Letters* **13** (1993), 9–12.

Interpretation of Dual Model for Piecewise Linear Programming Problem

Robert Hlavatý

Abstract. Piecewise linear programming models are suitable tools for solving situations of non-linear nature. Several approaches for solving such models have been developed and are already known. Depending on type of the objective function (convex or non-convex), there are, among others, two common ways of solving these models. The solution of the models can either be found by plain modification of simplex algorithm or, if non-convex, in a different specific ways. Using an auxiliary model of integer (bivalent) programming is one way to deal with the problem. For each linear model, there is also dual model and its properties and purpose is widely known.
This paper is to describe dual associated model for general piecewise linear model with auxiliary model. This dual model provides an additional knowledge useful for solution of models that cannot be solved in any common way. The properties of this dual model would be described further in this paper. Dual model solution optimality and feasibility are discussed.

Keywords: Piecewise linear programming, duality, simplex algorithm.

JEL Classification: C61
AMS Classification: 90C30

1 Introduction

Duality plays an important role in mathematical programming from both theoretical and computational point of view. [8] For those linear and convex problems very well known theory has been developed [3,12]. Studying duality for nonlinear programming problems has been of much interest in the past [7]. Various dual models have been developed. Dantzig et al. [4]. Mond [10] and Bazaraa and Goode [1] formulated a pair of symmetric dual models for nonlinear mixed integer programs. Mishra, Wang and Lai [9] followed that with their work, dealing with objective function convexity and concavity issues within such nonlinear models. Generally, more various specific models concerning duality of mathematical programming has been developed recently. Luc [8] dealt with duality in multiple objective linear programming, Fan [5] dealt with theories in nonlinear semidefinite programming and Mond and Weir [11] proposed a number of different duals for nonlinear programming problems with nonnegative variables and proved various duality theorems under appropriate pseudo-convexity/quasi-convexity assumptions.

There are many more approaches and models dealing with duality described by Mond and Chandra [2,10]. Difference between these duality models depends either on approach of defining and computing these models or on properties of original primal model. In this paper, let us focus on **piecewise linear model**. A general model of piecewise linear programming is defined in following way:

$$Z(x): \sum_{j=1}^{n} a_{ij}x_j \leq b_i, i = 1,2,\dots,m; x_j \geq 0, j = 1,2,\dots,n \tag{1}$$

Where $Z(x)$ is **objective function** consisting of number of fractional functions:

$$\begin{gathered} Z(x) = Z_1(x_1) + Z_2(x_2) + \cdots + Z_j(x_j) + \cdots + Z_n(x_n), \\ Z_j(x_j) = Z_{j1}(x_j), x_j \in \langle 0, k_{j1} \rangle \\ Z_j(x_j) = Z_{j1}(x_j), x_j \in \langle k_{j1}, k_{j2} \rangle \\ \vdots \\ Z_j(x_j) = Z_{jp_i}(x_j), x_j \in \langle k_{jp_i}, \infty \rangle, j = 1,2,\dots,n \end{gathered} \tag{2}$$

All of the functions Z_{jk} are considered to be linear.

The idea of this paper is to develop a dual model for specific primal model of nonlinear programming defined by Houška [6] called **general model of piecewise linear programming (PWLP)**. This particular model uses piecewise linear model with auxiliary model to solve problems of non-linear nature. Let us briefly describe this modified model and its components in the way Houška [6] suggested:

$$Z(x,r)\left|\begin{array}{c}\sum_{j=1}^{n} a_{ij}x_j \le b_i, i = 1,2,\dots,m; x_j \ge 0, j = 1,2,\dots,n \\ x_j + r_{j(2k-1)} - r_{j(2k)} = k_{jk}, j = 1,2,\dots,n, k = 1,2,\dots,p_j \\ r_{jk} \ge 0, j = 1,2,\dots,n; k = 1,2,\dots,p_j \\ r_{j(2k-1)}.r_{j(2k)} = 0, j = 1,2,\dots,n; k = 1,2,\dots,p_j\end{array}\right. \tag{3}$$

Where the linear **objective function** is now following:

$$Z(x,r) = \sum_{j=1}^{n} c_j x_j + \sum_{k=1}^{p_j} c_{jk} r_{jk}, j = 1,2,\dots,n \tag{4}$$

Modified model obviously contains the **original constraints**:

$$\sum_{j=1}^{n} a_{ij}x_j \le b_i, i = 1,2,\dots,m$$

And additional **constraints of boundary points** with left distance variables $r_{j(2k-1)}$ and right distance variables $r_{j(2k)}$ regarding dividing points between subsequent linear functions in piecewise linear objective function:

$$x_j + r_{j(2k-1)} - r_{j(2k)} = k_{jk}, j = 1,2,\dots,n, k = 1,2,\dots,p_j \tag{5}$$

New **cost coefficients** are:

$$c_j \text{and } c_{jk} \tag{6}$$

Solution of the modified model still must remain **nonnegative**. This requires both constraints of original structure variables and constraints of distance variables:

$$x_j \ge 0, j = 1,2,\dots,n \tag{7}$$

$$r_{jk} \ge 0, j = 1,2,\dots,n; k = 1,2,\dots,p_j \tag{8}$$

Following **constraints containing nonlinear elements** can be omitted in the calculation itself, thus it has no effect on improving the value of the objective function:

$$r_{j(2k-1)}.r_{j(2k)} = 0, j = 1,2,\dots,n; k = 1,2,\dots,p_j \tag{9}$$

To use the standard simplex algorithm for solving this model, it is vital that the **objective function is convex** and following rules concerning cost coefficients have to be fulfilled:

For maximization model:

$$c_j \ge c_j + c_{j2} \ge c_j + c_{j2} + c_{j4} \ge \cdots \ge c_j + \sum_{k=1}^{p_i} c_{jk} \tag{10}$$

For minimization model:

$$c_j \le c_j + c_{j2} \le c_j + c_{j2} + c_{j4} \le \cdots \le c_j + \sum_{k=1}^{p_i} c_{jk} \tag{11}$$

For this particular model of piecewise linear programming with auxiliary model, the dual model and its properties will be described further in this paper.

2 General dual model

2.1 Primal model modification necessity

With this primal model given, it is possible to create appropriate general dual model of piecewise linear programming. The procedure of creating will be similar to transforming a simple linear model to dual model. A matrix of constraint coefficients must be transposed for the dual model. Cost coefficients of the primal model will make right-hand side coefficients of the dual model. Right-hand side coefficients of the primal model will make cost coefficients of the dual model. Equalities and inequalities of conditions in the primal model will affect (depending whether the objective functions is max or min) conditions of non-negativity of the dual model. Conditions of non-negativity in the primal model will affect equalities or inequalities in the dual model conditions. Maximization objective function will turn to minimization and vice versa. It is, however, necessary to deal with several difficulties that occur while transferring to dual model, while the transformation of general PWLP model cannot be done so easily like for the simple linear model.

The rudimentary problem appears that in the primal model, there are multiple objective functions. As the right-hand side vector of the dual model is strictly dependent on the coefficients of primal objective function, it is uncertain which of these values should that be, as there are more than one of each in the primal model piecewise linear objective function. The right-hand side values of the dual model actually determine the set of feasible solutions. Existence of number of linear functions in the primal objective function leads to very complicated set of feasibility solutions. The higher the number of piecewise linear functions in the primal, the greater number of conditions is to be met in the dual model. For example a primal model with a piecewise linear objective function consisting of two linear functions, each with two dividing points, leads to six more conditions to be met in the dual model. Generally the number of these conditions is determined by all combinations between subsequent linear functions in the primal objective function. The higher number of such conditions would make extremely large dual model difficult both for calculation and interpretation.

2.2 Dealing with the non-negativity of dual model

It is then wiser to transform a primal model to (3) first and create a dual model afterwards. The problem with the numerous coefficients of the primal objective function is not present anymore, as there is only one objective function and the right-hand side coefficients of the dual model can be determined easily. However the primal model is now larger in the terms of constraints of boundary points (5). These constraints will always exist in the primal model in the form of equalities. The form of restrictive conditions of the primal model directly affects non-negativity conditions (this term is slightly improper for the dual model, as these conditions don't have to be generally non-negative) of the dual model. Because of these constraints of boundary points in the primal model there will always be conditions *sine limitis* (the value of the variable can be whatsoever negative, non-negative or zero). Using the standard simplex method requires all of the variables to be non-negative. To meet this requirement it is necessary to transfer *s.l. (sine limitis)* conditions of dual model in the following way:

Original constraints of dual model $y_n = s.l.$ have to be replaced with these two conditions:

$$y_n \geq b_m; y_n \leq b_m \tag{12}$$

These conditions have the same meaning (there are no limitations of y values), however the condition of non-negativity still have not been met because of the condition $y_n \leq 0$. This condition can be easily turned non-negative by a simple substitution:

$$y_n = -y_n' \tag{13}$$

A new condition is created:

$$y_n' \geq 0 \tag{14}$$

This condition meets the requirement of non-negativity. It is of course necessary to substitute given variable wherever it appears in the whole dual model. This actually means multiplying it by -1.

2.3 Preserving feasibility of dual model

As it was described earlier the right-hand side coefficients in dual depend on cost coefficients of objective function in primal. The primal cost coefficient can be often negative. This is due the decreasing slopes of piecewise linear functions within the maximization objective function (10). Also the structure variables costs could be negative for various reasons depending on an interpretation of the problem. Should there be one or more negative right-hand side coefficients in the dual model, it is necessary to change them to positive by plain multiplying the whole equality (or inequality) by -1. The primal feasibility is vital for using a standard simplex method for solving any model.

2.4 Definition of the dual model

After dealing with the several problems regarding feasibility of the dual model, it is now possible to define a general dual model for primal piecewise linear model. This model can exist in two forms depending on whether the objective function is maximization or minimization. In the general model description, primal variable is only described generally as x (for both x and r, because the way of transforming to dual model is the same for both of them).

Maximization primal model

Pairs of appropriate parts will exist in primal and dual models as follows (m-rows, n-columns):

Matrices of conditions coefficients:

$$\begin{matrix} a_{11}x_1 & a_{12}x_2 & \cdots & a_{1n}x_n \\ a_{21}x_1 & a_{22}x_2 & \cdots & a_{2n}x_n \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}x_1 & a_{m2}x_2 & \cdots & a_{mn}x_n \end{matrix} \rightarrow \begin{matrix} a_{11}y_1 & a_{21}y_2 & \cdots & a_{m1}y_m \\ a_{12}y_1 & a_{22}y_2 & \cdots & a_{m2}y_m \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n}y_1 & a_{2n}y_2 & \cdots & a_{nm}y_m \end{matrix} \tag{15}$$

Primal right-hand side vector and dual objective function:

$$\underset{b}{\rightarrow}(b_1, b_2, \dots, b_m)^T \quad \rightarrow \quad z(y) = b_1y_1 + b_2y_2 + \cdots b_my_m \rightarrow MIN \tag{16}$$

Primal cost coefficients and dual right-hand side vector:

$$z(x) = c_1x_1 + c_2x_2 + \cdots + c_nx_n \rightarrow MAX \quad \rightarrow \quad \underset{c}{\rightarrow}(c_1, c_2, \dots, c_n)^T \tag{17}$$

Primal non-negativity conditions and dual conditions symbols:

$$x_n \geq 0 \quad \rightarrow \quad \sum a_{mn}y_m \geq c_n \tag{18}$$

Primal conditions symbols and dual non-negative conditions symbols:

$$\sum a_{mn}x_n \leq b_m \quad \rightarrow \quad y_m \geq 0 \tag{19}$$

$$\sum a_{mn}x_n \geq b_m \quad \rightarrow \quad y_m = -y'_m;\ y'_m \geq 0 \tag{20}$$

$$\sum a_{mn}x_n = b_m \Rightarrow \sum a_{mn}x_n \leq b_m \wedge \sum a_{(m+1)n}x_n \geq b_m \quad \rightarrow \quad y_m \geq 0; \quad y_{m+1} = -y'_{m+1};\ y'_{m+1} \geq 0 \tag{21}$$

Minimization primal model

Pairs of appropriate parts will exist in primal and dual models as follows (m-rows, n-columns):

Matrices of conditions coefficients:

$$\begin{matrix} a_{11}x_1 & a_{12}x_2 & \cdots & a_{1n}x_n \\ a_{21}x_1 & a_{22}x_2 & \cdots & a_{2n}x_n \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}x_1 & a_{m2}x_2 & \cdots & a_{mn}x_n \end{matrix} \rightarrow \begin{matrix} a_{11}y_1 & a_{21}y_2 & \cdots & a_{m1}y_m \\ a_{12}y_1 & a_{22}y_2 & \cdots & a_{m2}y_m \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n}y_1 & a_{2n}y_2 & \cdots & a_{nm}y_m \end{matrix} \tag{22}$$

Primal right-hand side vector and dual objective function:

$$\underset{b}{\rightarrow}(b_1, b_2, \dots, b_m)^T \quad \rightarrow \quad z(y) = b_1y_1 + b_2y_2 + \cdots b_my_m \rightarrow MAX \tag{23}$$

Primal cost coefficients and dual right-hand side vector:

$$z(x) = c_1x_1 + c_2x_2 + \cdots + c_nx_n \rightarrow MIN \quad \rightarrow \quad \underset{c}{\rightarrow}(c_1, c_2, \dots, c_n)^T \tag{24}$$

Primal non-negativity conditions and dual conditions symbols:

$$x_n \geq 0 \quad \rightarrow \quad \sum a_{mn}y_m \leq c_n \Rightarrow -\sum a_{mn}y_m \geq c_n \tag{25}$$

Primal conditions symbols and dual non-negative conditions symbols:

$$\sum a_{mn}x_n \geq b_m \quad \rightarrow \quad y_m \geq 0 \tag{26}$$

$$\sum a_{mn}x_n \leq b_m \quad \rightarrow \quad y_m = -y'_m;\ y'_m \geq 0 \tag{27}$$

$$\sum a_{mn}x_n = b_m \Rightarrow \sum a_{mn}x_n \geq b_m \wedge \sum a_{(m+1)n}x_n \leq b_m \quad \rightarrow \quad y_m \geq 0; \quad y_{m+1} = -y'_{m+1};\ y'_{m+1} \geq 0 \tag{28}$$

3 Illustrative example

Let us have following random **primal model of piecewise linear programming** containing 3 limiting conditions, non-negative conditions and piecewise linear function consisting of 6 linear sub-functions:

$$x_1 + x_2 \leq 24\ ; \quad x_1 + 0{,}5x_2 \leq 16\ ;\ x_1 \geq 10\ ;\ x_1, x_2 \geq 0$$

$$z = Z(x_1) + Z(x_2) \rightarrow MAX$$

$$Z(x_1) = 1{,}5x_1 \in (0\ ,2), Z(x_1) = x_1 + 1 \in (2\ ,3), Z(x_1) = 0{,}5x_1 + 2{,}5\ (3\ ,\infty)$$

$$Z(x_2) = 2x_2 \in (0{,}1), Z(x_2) = x_2 + 1 \in (1\ ,3{,}5), Z(x_2) = 0{,}6x_1 + 2{,}4\ (3{,}5\ ,\infty)$$

Transferring it to model (3) additional **boundary points conditions** are added:

$$x_1 + r_{11}^- - r_{11}^+ = 2\ ;\ x_1 + r_{12}^- - r_{12}^+ = 3\ ;\ x_2 + r_{21}^- - r_{21}^+ = 1\ ;\ x_2 + r_{22}^- - r_{22}^+ = 3{,}5$$

And appropriate objective function is constructed:

$$z = 1{,}5x_1 - 2x_2 - 0{,}5r_{11}^+ - 0{,}5r_{12}^+ - r_{21}^+ - 0{,}4r_{22}^+ \rightarrow MAX$$

Using the standard simplex algorithm the optimal solution of this model can be found. Expressing it as a base solution vector the **solution is following**:

$$x_B^T = (10\ ,12\ ,0\ ,8\ ,0\ ,7\ ,0\ ,11\ ,0\ ,8{,}5)$$

Optimal value of the objective function z is 17,1.

Let us find a dual model of the given primal model and find a solution of this dual model using a general model (15-28) described in the previous chapter. All necessary modifications preserving model feasibility and non-negativity has been made immediately in one step, resulting in the following model:

$$y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 \geq 1{,}5\ ;\ y_1 + 0{,}5y_2 + y_8 - y_9 + y_{10} - y_{11} \geq 2$$

$$y_4 - y_5 \geq 0\ ;\ y_4 - y_5 \leq 0{,}5\ ;\ y_6 - y_7 \geq 0\ ;\ y_6 - y_7 \leq 0{,}5$$

$$y_8 - y_9 \geq 0\ ;\ y_8 - y_9 \leq 1\ ;\ y_{10} - y_{11} \geq 0\ ;\ y_{10} - y_{11} \leq 0{,}4$$

$$y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10}, y_{11} \geq 0$$

$$z(y) = 24y_1 + 16y_2 - 10y_3 + 2y_4 - 2y_5 + 3y_6 - 3y_7 + y_8 - y_9 + 3{,}5y_{10} - 3{,}5y_{11} \rightarrow MIN$$

If there is an existing optimal solution with the finite value of the objective function in the primal model, then there also has to be the same for the dual model. Using the standard simplex algorithm, the optimal solution of the dual model has been found. The **value of the objective function is again 17,1**. Because it is dual model the

interpretation of itself is different from primal model. Optimal variable values of primal model can be found in the dual model as well, though they exist here as the shadow costs and in the optimal simplex table they can be found in optimality test row under the slack variables of dual model. Also their values are opposite (.-1), because it is necessary to have only negative values in the last row of the simplex table. Should the primal model be minimization and dual model maximization, then the optimal values in the dual simplex table would be the same as in primal simplex table. Generally, the appropriate pairs of values will be located in both simplex tables in the following way:

Primal model	Dual model
Optimal values of structure variables	Shadow costs of slack variables
Optimal values of slack variables	Shadow costs of structure variables
Dual costs of structure variables	Optimal values of slack variables
Dual costs of slack variables	Optimal values of structure variables

Table 1 Relations between primal and dual optimal solutions

4 Conclusion

The general dual model for primal piecewise linear programming model has been found and described. If one of these associated models have an optimal solution with a finite value of objective function then the second model also have an optimal solution with a finite value of objective function. Both values of objective function must be equal. Knowing optimal solutions of both associated problems, the dual values (dual costs) represent a change of the value of primary objective function upon unit increase of right-hand sides of conditions in primal problem. Optimal solution (optimal program) can exist if there is also an optimal valuation of factors. Dual model of piecewise linear programming makes it possible to do further interpretation of problems of non-linear nature, which can be primarily solved by primal model. Finding the solution of dual model brings other possibilities of analysis for given economical problem.

Acknowledgements

This paper was created under the project Solution of nonconvex models of piecewise linear programming, No.20121055, with the support of Internal Grant Agency, Czech University of Life Sciences.

References

[1] Bazaraa, M.S., and Goode, J.J.: On symmetric duality in nonlinear programming, Operations Research 21/1 (1973) 1-9.
[2] Chandra, S., Abha: A note on pseudo-invexity and duality in nonlinear programming. European Journal of Operational Research 2000, 122, 161–165.
[3] Dantzig, G.B.: Linear Programming and Extension, Princeton, New Jersey, 1963.
[4] Dantzig, G.B.: Eisenberg, E., and Cottle, R.W., Symmetric dual nonlinear programming, Pacific Journal of Mathematics 15 (1965) 809-812.
[5] Fan, J.: Duality theories in nonlinear semidefinite programming. Appl. Math. Lett. 2005, 18, 1068–1073.
[6] Houška, M., Beránková, M.: Sensitivity Analysis in Piecewise Linear Programming Models. In: Proceedings of the 23rd International Conference Mathematical Methods in Economics 2005, Hradec Králové: 2005, pp.140 -145.
[7] Kumar, V., Husain, I., Chandra, S.: Symetric duality for minimax nonlinear mixed integer programming. Europian Journal of Operational Research 1995, 80, 425–430.
[8] Luc, D. T.: On duality in multiple linear objective programming. Europian Journal of Operational Research 2011, 210, 158–168.
[9] Mishra, S. K., Wang, S. I., Lai, K. K.: Symmetric duality for minimax mixed integer programming problems with pseudo-invexity. Europian Journal of Operational Research 2009, 198, 37–42.
[10] Mond, B.: A symmetric dual theorem for nonlinear programs, Quarterly Journal of Applied Mathematics 23 (1965) 265-269.
[11] Mond, B., Weir, T.: Generalized concavity and duality, in: S.Schaible, W.T. Ziemba (Eds.), Generalized Concavity in Optimization and Economics, Academic Press, New York,1981, pp. 263–279.
[12] Rockafellar, R.T.: Convex Analysis, Princeton, New Jersey, 1970.

DSGE model with collateral constraint: estimation on Czech data

Miroslav Hloušek[1]

Abstract. Czech data shows positive comovement of house prices and consumption in reaction to house price shock. This behavior can be explained by collateral effect when houses serve as collateral for credit constrained households. This type of friction is present in the Dynamic Stochastic General Equilibrium (DSGE) model from Iacoviello [3] which is slightly modified and estimated on Czech data using Bayesian techniques. The estimated parameters are economically interpreted and ability of the model to match moments in data is assessed. Situation when houses are not collateralizable is examined. This exercise shows that the collateral effect is necessary feature of the model to deliver positive reaction of consumption to house price shock.

Keywords: collateral constraint, housing, DSGE model, Bayesian estimation.

JEL classification: E370
AMS classification: 91B64

1 Introduction

It is empirical fact that consumption and house prices comove over the business cycle. This is true also in the Czech economy. Looking at correlation between consumption and house prices (both variables are expressed in gaps), we get quite large value of correlation coefficient, $\rho_{c,q} = 0.68$, for output and house prices it is somewhat smaller, $\rho_{y,q} = 0.35$. The tight relationship between house prices and consumption is confirmed by structural VAR model. Figure 1 shows reaction of house prices, q_t, consumption, C_t, output, Y_t, and interest rate, R_t, to house price shock.[2] There is evident positive comovement of consumption and output with house prices in response to house price shock.

This empirical fact can be explained by existence of collateral effect – mechanism incorporated in presented model. The model is taken from Iacoviello [3] and includes credit constrained households (and firms) which need to collateralize their loans. The mechanism closely follows Kiyotaki and Moore [5], but instead of land, houses serve as collateral. Next feature is that the debt is quoted in nominal terms which is based on empirical grounds from low-inflation countries. This makes another channel for propagation of financial shocks into real part of economy. The transition mechanism is as follows: positive demand shock increases price of assets (housing) which increases borrowing capacity of constrained households/firms and allows them to spend and invest more. The rise in prices reduces the real value of their debt obligations, which further increases value of their net worth. Borrowers have higher propensity to spend than lenders and thus the net demand is positively affected. This mechanism works as amplification of demand shocks. However, last mentioned price effect also works for supply shocks (which are characterized by negative correlation between output and prices). In case of adverse supply shocks, this mechanism helps to restore long run equilibrium, because it supports spending and investing. Thus there is accelerator of demand shocks and decelerator of supply shocks. However, in both cases the model predicts positive relationship between house prices and consumption. Given recent developments at housing market, it seems quite important to understand this mechanism and to verify it on Czech data.

The rest of the paper is organized as follows. Section 2 presents main parts of the model, Section 3 briefly describes data and estimation technique. Results of the estimation, data fit of the model and dynamical properties of alternative settings are discussed in Section 4. Final section concludes with prospects for further research.

[1]Masaryk University, Department of Economics, Lipová 41a, 602 00 Brno, hlousek@econ.muni.cz

[2]The model was estimated on Czech data spanning from 1998Q1 to 2011Q3; ordering of variables is following: R_t, q_t, C_t, Y_t, π_t. More information on data is in section 3.

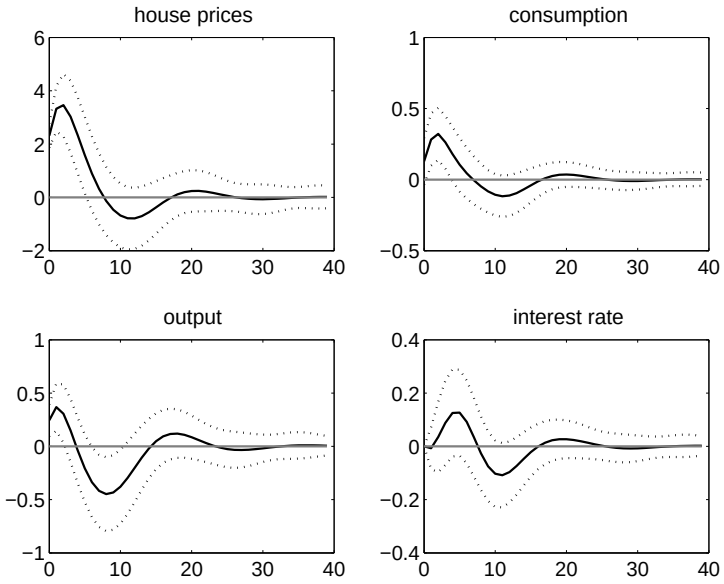


Figure 1 VAR evidence: Impulse responses to house price shock

2 Model

2.1 Households

The model is borrowed from Iacoviello [3] and slightly adjusted, especially for estimation purposes. It includes two types of households: patient and impatient (indexed by $i = 1,2$). They differ by the time discount factor, $\beta_i \in (0,1)$, where $\beta_1 > \beta_2$, i.e. impatient households has lower discount factor and thus discounts future more heavily. Both households consume $c_{i,t}$, supply labor $L_{i,t}$, accumulate housing $h_{i,t}$ and real money balances $M_{i,t}/P_t$ where M_t are nominal balances and P_t denotes price level in time t. They maximize utility function

$$E_0 \sum_{t=0}^{\infty} \beta_i^t \left(\ln c_{i,t} + j_t \ln h_{i,t} - \frac{(L_{i,t})^\eta}{\eta} + \chi \ln \frac{M_{i,t}}{P_t} \right)$$

where j_t is housing demand shock which can be also interpreted as shock to house prices, η denotes slope of labor supply and χ is weight to money holdings. The budget constraint is

$$c_{i,t} + q_t \Delta h_{i,t} + \frac{R_{t-1} b_{i,t-1}}{\pi_t} = b_{i,t} + w_{i,t} L_{i,t} + F_t + T_{i,t} - \Delta \frac{M_{i,t}}{P_t}$$

where $\pi_t = \frac{P_t}{P_{t-1}}$ is inflation, $q_t = \frac{Q_t}{P_t}$ is the real housing price, $w_{i,t} = \frac{W_{i,t}}{P_t}$ is the real wage and $b_{i,t} = \frac{B_{i,t}}{P_t}$ denotes loans in real terms. F_t are lump-sum profits from retailers that goes only to patient households, and $T_{i,t} - \Delta \frac{M_{i,t}}{P_t}$ are net transfers from the central bank. The term $\frac{R_{t-1}}{\pi_t}$ reflects the assumption that debt contracts are set in nominal terms. Changes in prices between $t-1$ and t thus can affect the realized real interest rate.

Unlike patient households, the impatient households are credit constrained. The maximum amount $B_{2,t}$ they can borrow (in nominal terms) is $m_h E_t \frac{Q_{t+1} h_{1,t}}{R_t}$. In the real terms:

$$b_{2,t} \leq m_h E_t \frac{q_{t+1} h_{2,t}}{R_t / \pi_{t+1}} \tag{1}$$

where m_h is loan-to-value ratio, i.e. the limit for borrowing expressed as the fraction of asset value (house). If the borrowers fail to repay their debt, the lenders can repossess the assets (housing), but must pay proportional transaction cost $(1-m_h)E_t(q_{t+1}h_{1,t})$. Under reasonable assumption, in the steady-state and in its neighborhood (given some uncertainty) the borrowing constraint (1) will hold with equality.

2.2 Entrepreneurs

Entrepreneurs produce intermediate goods Y_t according to Cobb-Douglas production function

$$Y_t = A_t K_{t-1}^{\mu} h_{t-1}^{\nu} L_{1,t}^{\alpha(1-\mu-\nu)} L_{2,t}^{(1-\alpha)(1-\mu-\nu)}$$

where A_t is shock to productivity, $L_{1,t}$ and $L_{2,t}$ are hours of work supplied by patient and impatient households, K_t is capital that is created at the end of each period. Entrepreneurs also consume and maximize utility function

$$E_0 \sum_{t=0}^{\infty} \gamma^t \ln c_t,$$

where γ is discount factor, subject to flow of funds

$$\frac{Y_t}{X_t} + b_t = c_t + q_t \Delta h_t + b_{t-1} \frac{R_{t-1}}{\pi_t} + w_{i,t} L_{i,t} + I_t + \xi_{K,t}$$

where I_t is the investment that follows from law of motion for capital, $K_t = (1-\delta)K_{t-1} + i_t I_t$, where i_t is investment efficiency shock. The term $\xi_{K,t}$ denotes capital adjustment cost, $\xi_{K,t} = \psi (I_t/K_{t-1} - \delta)^2 K_{t-1}/2\delta$. Similarly to impatient households, entrepreneurs have lower discount factor, $\gamma < \beta_1$, and are credit constrained

$$b_t \leq m_e E_t \frac{q_{t+1} h_t}{R_t/\pi_{t+1}}$$

where m_e denotes loan-to-value ratio. Again the borrowing constraint is binding around the steady state.

2.3 Retailers and monetary authority

Retailers are incorporated in the model only for the sake of introducing nominal rigidity. They operate at monopolistically competitive market. They purchase the intermediate good from entrepreneurs at the wholesale price P_t^w, transform it into composite final good and sell at price P_t with markup $X_t = \frac{P_t}{P_t^w}$. The price setting is modeled in Calvo [2] style. Their optimization problem is quite standard and leads to conventional New Keynesian Phillips curve which has following log-linearized form: [3]

$$\hat{\pi}_t = \beta E_t \hat{\pi}_{t+1} - \kappa \hat{X}_t + \hat{u}_t$$

where $\kappa = \frac{(1-\theta)(1-\beta\theta)}{\theta}$, θ is Calvo parameter (probablity of not resetting the price), $\hat{X}_t$ is the deviation of markup from steady state and $\hat{u}_t$ is cost-push shock.

The central bank behaves according to Taylor rule with interest rate smoothing (in log-linearized form):

$$\hat{R}_t = r_R \hat{R}_{t-1} + (1-r_R)[(1+r_\pi)\hat{\pi}_t + r_Y \hat{y}_t] + \hat{e}_{R,t}$$

where and $\hat{e}_{R,t}$ is shock to monetary policy which is assumed *iid* with zero mean and variance σ_R^2.

2.4 Equilibrium

There is unique stationary equilibrium, entrepreneurs and impatient households hit the borrowing constraint, borrow up to the limit, make the interest payments on the debt and roll the steady state stock of debt over forever. Markets for labor, housing, goods and loans clear. For estimation purposes, stochastic shock, e_Y, is added to market clearing condition for goods market $Y_t = c_t + c_{1,t} + c_{2,t} + I_t + e_Y$. It should capture other effects such as government expenditures or net exports and bring the model closer to Czech data. This shock, e_Y, and monetary policy shock, e_R, are assumed *iid* processes, shocks to technology A_t, housing preferences j_t, cost-push shocks u_t and investment shocks i_t follow AR(1) processes.

The steady state of the model is derived and model equations are log-linearized around it. The model is transformed into state space system and solved using Klein [6] procedure.

[3] The variables with hat are expressed as deviation from steady state.

3 Data and estimation

The model is estimated using data for following model variables: output (Y_t), consumption (C_t), investment (I_t), real house prices (q_t), inflation (π_t) and nominal interest rate (R_t). Time series are quarterly, they are obtained from the Czech Statistical Office and the Czech National Bank and cover time period 1998:Q1 – 2011:Q3. Specifically, output is gross domestic product (GDP), investment is gross fixed capital formation, consumption is measured by expenditure of households, interest rate is represented by 3M Pribor, inflation rate is q-on-q change of consumer price index (CPI) and real house prices are represented by index of realized (offering) prices of flats deflated with CPI. Data for output, consumption and investment are expressed in per capita terms. Data for output, investment, consumption, real house prices and nominal interest rate are detrended using Hodrick-Prescott filter (with λ = 1600). Inflation is demeaned and annualized.

Some of the model parameters are calibrated according to Iacoviello [3] and data from national accounts. Description of calibrated parameters and their values are quoted in Table 1. The rest of the model parameters is then estimated using Bayesian techniques. It combines maximal-likelihood with some prior information to get posterior distribution of the parameters. Specifically, posterior inference was obtained by Random Walk Chain Metropolis-Hastings algorithm which generated 1000,000 draws from the posterior distribution. They were computed in two chains with 500,000 replications each, 50 % 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.

Description	Param.	Value	Description	Param.	Value
Preferences			*Technology*		
Discount factor: Patient HH	β_1	0.99	Calvo parameter	θ	0.75
Discount factor: Impartient HH	β_2	0.95	Capital share	μ	0.30
Discount factor: Entrepreneurs	γ	0.98	Housing share	ν	0.05
Labor supply aversion	η	1.01	Capital depreciation rate	δ	0.05
Weight on housing	j	0.10	Steady-state markup	X	1.10

Table 1 Calibrated parameters

4 Results of estimation

The prior means and standard deviations of estimated parameters are quoted in Table 2. The priors are set according to Iacoviello [3]. Table 2 also shows the posterior means of estimated parameters together with 95 % confidence intervals. The labor share of patient households, α, is 0.46. It is lower than the prior and also lower than values found in other empirical studies for U.S. economy or Canada (see Iacoviello and Neri [4] or Christensen et al. [7]). This value implies that the share of borrowing constrained households ($1-\alpha = 0.54$) in the Czech economy is larger than that of unconstrained, which should contribute to positive elasticity of consumption to house prices. Loan-to-value ratios for entrepreneurs and impatient households are $m_e = 0.51$ and $m_h = 0.79$, respectively. It means that houses owned by impatient households are more easily collateralizable than entrepreneurial real estates. This result differs from Iacoviello [3] who found the opposite on U.S. data. Posterior mean of ψ is 2.39 and shows quite high adjustment cost of investment. Parameters of monetary policy rule are quite standard and correspond to other empirical studies for the Czech economy. Regarding the shocks, the most persistent is the shock to housing preferences ($\rho_j = 0.94$), the least persistent is technology shock ($\rho_A = 0.59$). The most volatile shock is also housing preference shock, with standard deviation $\sigma_j = 26.28$. It is quite intuitive because the examined period includes house price boom in 2008 and subsequent decline. The house prices fluctuated a lot and it is something than cannot be explained by model itself.

Next step is evaluation of data fit of the model. Table 3 shows moments calculated from data and moments obtained from model simulations (with 90 % confidence bands). The outcome of the model is quite poor. The volatility of output and consumption in model is much higher than in data while volatility of inflation is lower than in data. On the other hand the volatility of investment, real house prices and interest rate is matched quite precisely. Relative volatilities (to output) implied by the model are also not satisfactory. E.g. investment is less volatile than output which contradict the data and consumption has almost same volatility as output. The model is able to match positive correlations between output and consumption, investment, and the real house prices but it fails to replicate the magnitude of correlations. The model has also problems to generate positive correlation between output and

	Prior distribution			Posterior distribution		
Parameter	Density	Mean	S.D.	Mean	2.5 %	97.5 %
Production						
α	beta	0.600	0.10	0.4586	0.2938	0.6243
ψ	normal	2.000	0.50	2.9487	2.2953	3.5930
LTV ratios						
m_e	beta	0.800	0.10	0.5119	0.3651	0.6612
m_h	beta	0.800	0.10	0.7938	0.7042	0.8869
MP rule						
r_R	normal	0.800	0.10	0.8076	0.7641	0.8514
r_π	normal	0.600	0.10	0.6140	0.4564	0.7699
r_Y	normal	0.125	0.10	0.0671	0.0266	0.1082
Persistence of shocks						
ρ_u	normal	0.700	0.10	0.7625	0.6945	0.8324
ρ_j	normal	0.700	0.10	0.9436	0.9189	0.9696
ρ_A	normal	0.700	0.10	0.5896	0.4924	0.6884
ρ_I	normal	0.700	0.10	0.7543	0.6121	0.9017
Volatility of shocks						
σ_R	normal	0.100	inf	0.0086	0.0063	0.0108
σ_u	normal	0.100	inf	0.0142	0.0102	0.0180
σ_j	normal	0.100	inf	0.2628	0.1645	0.3592
σ_A	normal	0.100	inf	0.0668	0.0486	0.0848
σ_I	normal	0.100	inf	0.0070	0.0031	0.0107
σ_Y	normal	0.001	inf	0.0139	0.0117	0.0162

Table 2 Prior and posterior distribution of structural parameters

inflation and output and interest rate found in data. On the other hand, the key positive correlation between the real house prices and consumption is captured almost precisely by the model.

	Data	Model				Data	Model		
		Mean	5 %	95 %			Mean	5 %	95 %
Volatility					Correlations				
Y	2.05	7.43	5.58	9.71	Y,C	0.52	0.99	0.99	1.00
C	1.12	6.60	4.99	8.76	Y,I	0.77	0.92	0.86	0.97
I	4.71	4.63	3.47	6.32	Y,q	0.35	0.69	0.49	0.85
q	7.69	6.61	4.85	8.91	Y,R	0.49	-0.73	-0.84	-0.51
R	1.18	1.37	1.05	1.69	Y,π	0.41	-0.05	-0.29	0.17
π	3.58	2.45	1.90	2.83	C,q	0.69	0.67	0.44	0.85

Table 3 Moments from data and model

The main message of the paper is to show the importance of collateral effect in generating positive comovement of house prices and consumption (output) in reaction to house price shock. This is examined by comparison of impulse response functions for two cases: model with all estimated parameters and model with parameters m_e and m_h set to 0. This can be interpreted as the houses (real estates) are not collateralizable at all and thus entrepreneurs and impatient households are excluded from financial markets. Result of this exercise is shown in Figure 2, which depicts reaction of key variables to house price shock. For the benchmark model, there is positive comovement of house prices and aggregate consumption (and output) which is also present in data as was shown in Figure 1. On the other hand, when collateral effect is shut down, initial reaction of consumption is negative. Subsequent positive deviation is negligible and the overall response is at odds with data.

5 Conclusion

This paper presented results of estimation of model with collateral constraint. The model successfully replicates empirical fact found in Czech data: positive correlation between consumption and house prices and positive co-movement of these two variables in reaction to house price shock. Even if the model captures this relationship, it fails in some other aspects. Volatility of several model variables is not in accordance with data and also some correlations differ in the magnitude and even the sign. The reason can be that the model omits some important

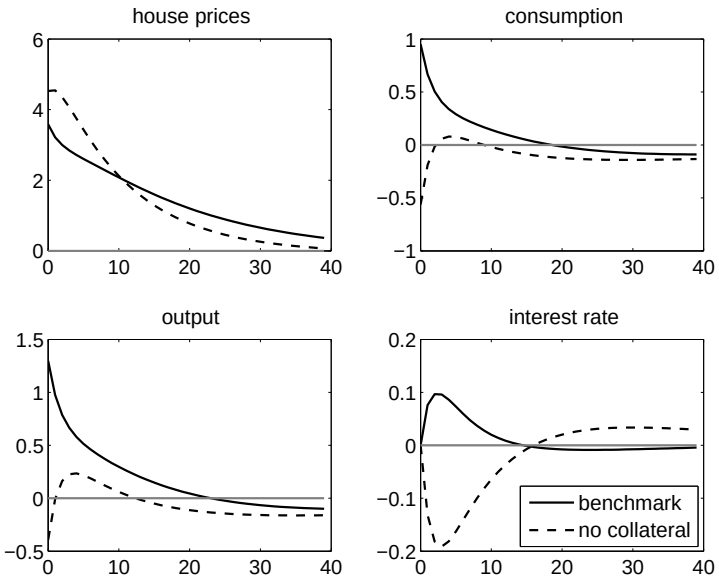


Figure 2 Impulse responses to house price shock

channels. It is closed economy model and includes only one type of nominal rigidity. The topic for further research is extension of the model by foreign sector and other nominal and real rigidities.

Acknowledgements

This work is supported by the Research Centre for the Competitiveness of the Czech Economy at Masaryk University and B5/11 research project of the CNB. The access to Wieland et al. [8] model database is highly appreciated.

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] Calvo, G.: Staggered prices in a utility-maximizing framework, *Journal of Monetary Economics 12*, 383–398, 1983.

[3] Iacoviello, M.: House Prices, Borrowing Constraints, and Monetary Policy in the Business Cycle, *American Economic Review*, Vol. 95, No. 3, pp. 739-764, 2005.

[4] Iacoviello, M. and Neri, S.: Housing Market Spillovers: Evidence from an Estimated DSGE Model, *American Economic Journal: Macroeconomics*, Vol. 2, No. 2, pp. 125Ű164, 2010.

[5] Kiyotaki, N. and Moore, J.: Credit Cycles, *The Journal of Political Economy*, Vol. 105, No. 2., pp. 211-248, 1997.

[6] Klein, P.: Using the generalized Schur form to solve a multivariate linear rational expectations model, *Journal of Economic Dynamics and Control*, Vol. 24, Issue 10, pp. 1405-1423, 2000.

[7] Christensen, I., Corrigan, P., Mendicino, C. and Nishiyama, S.: Consumption, Housing Collateral, and the Canadian Business Cycle, Bank of Canada, *Working paper*, 2009-26, 2009.

[8] Wieland, V., Cwik, T., Müller, G. J., Schmidt, S. and Wolters, M. *A New Comparative Approach to Macroeconomic Modeling and Policy Analysis*, Working Paper, Goethe University of Frankfurt, May 2011 (revised version).

Quantitative measuring of operational complexity of supplier-customer system with control thresholds

Jiří Hofman [1], Ladislav Lukáš [2]

Abstract. Measurement of operational complexity of company supplier-customer relations belongs to important managerial tasks. In the present work, we deal with quantitative measuring of these relations on the base of entropy approach. We assume a problem-oriented database exists, which contains detailed records of all product orders, deliveries and forecasts both in quantity and time being scheduled and realized, within time period given. However, managerial practice shows that instead of crisp definition of order and delivery variations very often more flexible approach based upon control thresholds is utilized. Application of such thresholds provides sensitive instruments to evaluate conditional entropy of supplier-customer relations. General procedure consists of three basic steps – pre-processing of data with consistency checks in Java, calculation of histograms and empirical distribution functions conditioned by control thresholds applied, and finally, evaluation of conditional entropy, which are realized both by Mathematica modules. Illustrative results of supplier-customer system analysis with and without time and volume quantity variation thresholds from selected Czech SME are presented in detail.

Keywords: Business economics, supplier-customer systems, firm performance, complexity measures, information and entropy.

JEL Classification: C63, C81, L25, M21.
AMS Classification: 91B42

1 Introduction

Business economics knows two types of complexity of supplier-customer systems, a structural complexity and an operational one, in principle. As usual, the structural complexity is defined as static variety of system and their design dimensions, and it describes links among various business units and their hierarchies. It has dominantly a static representation and undergoes time changes usually in long-term periods.

On the contrary, the operational complexity can be defined by uncertainties associated with dynamics of system. Hence, it reflects temporal changes in supplier-customer system, and an operational complexity measure should express behavioral uncertainties of the system during the time with respect to specified levels of control. We know that operational complexity of supplier-customer system is associated with specific data provided by inventory management. It has to record all possible types of flow variations within and across companies in detail, e.g. replenishment time disturbances, deviations of material in/out flows, etc.

2 Theoretical background

The theoretical framework for quantification of any system complexity is provided by information theory, in general, and we may refer [4] for more details. The Shannon's information-theoretic measure and corresponding entropy are the most important quantitative measures of expected amount of information required to describe the state of a system. In general, the complexity of a system increases with increasing levels of disorder and uncertainty of its states.

Basic mathematical model of information complexity assumes that N objects are given. The procedure follows an idea that as any object has to be identified uniquely as unique binary code $(a_1,..., a_d)$ has to be assigned to each object, where a_i , i=1,..., d belongs to {0,1}, where d denotes the highest exponent satisfying the relation $2^d \leq N$. Hence, the integer d satisfies $0 \leq d - \log_2 N < 1$. It is well-known that quantity $I = \log_2 N$ gives the length of most effective binary coding for unique identification of N objects.

In probabilistic framework, a random trial issues an event A_i, which belongs to given complete set of mutually disjunctive events $\{A_1,...,A_N\}$ having probabilities $p_i = P(A_i)$, i=1,..., N and satisfying $p_1+...+p_N = 1$. Making a large number of independent trials n, we get ratios $n(A_i)/n$ approaching p_i, i=1,..., N, where $n(A_i)$ denotes the number of occurrences of event A_i within the n independent trials. There is evident that $n(A_1)+...+ n(A_N) = n$

[1] University of West Bohemia, Faculty of Econonics, Husova 11, 30614 Plzeň, hofman@kpm.zcu.cz .
[2] University of West Bohemia, Faculty of Econonics, Husova 11, 30614 Plzeň, lukasl@kem.zcu.cz.

holds. The total number of possible outcomes, where events A_i, i=1,..., N, appear $n(A_i)$ times each, is $N_n = n!/(n_1!... n_N!)$, with $n_i \approx np_i$. However, to calculate factorials for large numbers is both tedious and error-prone numerical problem, so the well-known Stirling asymptotic formula is ready to use expressing $\log_2(N_n)$ in analytic form, for $n \to \infty$.

$$m! \approx m^m e^{-m} \sqrt{(2\pi m)}, \text{ for large integer } m.$$

After some technical manipulation following [4] we get

$$\log_2(N_n) \approx n\log_2(n) - \sum_{i=1}^{n} np_i \log_2(np_i) + \log_2(\sqrt{(2\pi n)} - \sum_{i=1}^{n} \log_2(\sqrt{(2\pi n_i)})) = - n \sum_{i=1}^{n} p_i \log_2(p_i) \quad (1)$$

Using (1) together with binary coding principle we get quantity d_n to be an estimation of the most effective binary coding length of any outcome from all possible N_n ones. It takes the following form

$$d_n \approx \log_2(N_n) \approx - n \sum_{i=1}^{n} p_i \log_2(p_i). \quad (2)$$

Now, having n independent trials, (2) provides a definition of new quantity I expressing the most effective binary coding for each individual trial in average

$$I = - \sum_{i=1}^{n} p_i \log_2(p_i). \quad (3)$$

Finally, we know that (3) provides an introduction of another quantity $I(p_1,..., p_N)$, which gives an average of information gauging appearance of events $\{A_1,..., A_N\}$ carried in any individual trial

$$I(p_1,..., p_N) = - \sum_{i=1}^{N} p_i \log_2(p_i). \quad (4)$$

However, the quantity $I(p_1,..., p_N)$ depends upon particular event probability distribution. The worst case on uncertainty is represented by random variable having uniform distribution. Let us denote this quantity I_u and using (4) we may calculate it in direct way

$$I(p_1,..., p_N) = - \sum_{i=1}^{N} p_i \log_2(p_i) = I_u = - \sum_{i=1}^{N} (1/N) \log_2(1/N) = \log_2(N), \quad (5)$$

The information-theoretic measure defined by expression (4) of a system, which states are completely described by mutually disjoint events $\{A_1,..., A_N\}$ with probabilities $(p_1,..., p_N)$, is called *entropy* of the system.

3 Operational complexity of supplier-customer system

In order to analyze operational complexity of supplier-customer system we have to define basic entities and corresponding flows of information. In principle, we are able to identify three main entities – supplier, customer, and an interface between them. In general framework, supplier and customer both schedule and realize their productions. Thus, the operational complexity is measurable at the interface where forecast, order and actual deliveries are detected.

Operational complexity of supplier-customer system results at this interface where the actual deliveries of goods deviate in quantity and/or time from that expected ones. Hence, the desired quantity will measure the amount of information required to describe the state of system in terms of the quantity and time variations across material flows and time-information flows that exist. From management science point of view, any supplier-customer system belongs to inventory control theory.

In order to get a structure of supplier-customer system which fits into a framework of information theory, we have to define both types of flows and set of events $\{A_1,..., A_N\}$ with their probabilities $(p_1,..., p_N)$ describing the system states. In general, the first task to identify is much simpler than the second one which needs a lot of problem-oriented data to be monitored, stored and evaluated.

Definition of variables.
We consider a set of products $\{P_1,..., P_n\}$ within a supplier-customer system. For any product P_i, we may consider two types of variables relating quantity and time in general, which provide time series being monitored. Notations proposed for these variables should consider both entities and production phases. Following [1, Ch.7], we introduce set of quantities denoted ${}_{(e,r)}Q_i$, and ${}_{(e,r)}T_i$, i=1,..., n, where e stands for entity, i.e. s for supplier, i for interface, and c for customer, and e stands for production phases, i.e. s for scheduled, p for actual production, and f, o, d for forecast, order and delivery, respectively. Their values should be kept in problem-oriented database.

Having such problem-oriented database available we may define proper set of events $\{A_1,..., A_N\}$ and calculate their probabilities $(p_1,..., p_N)$. We have to note that this step is much more involved and may depend on various aspects of supplier-customer system investigated, in particular upon its structural complexity, too. We assume the quantities ${}_{(e,r)}Q_i$, and ${}_{(e,r)}T_i$, i=1,..., n, to be represented by real continuous scalar variables. Then, each event A_k is defined by flow variation considered, e.g. (Order – Forecast), (Delivery – Order) and (Actual production – Scheduled production)

$({}_{i,o}Q_i - {}_{i,f}Q_i)$, $({}_{i,o}T_i - {}_{i,f}T_i)$, $({}_{i,d}Q_i - {}_{i,o}Q_i)$, $({}_{i,d}T_i - {}_{i,o}T_i)$,
$({}_{s,p}Q_i - {}_{s,s}Q_i)$, $({}_{s,p}T_i - {}_{s,s}T_i)$, $({}_{c,p}Q_i - {}_{c,s}Q_i)$, $({}_{c,p}T_i - {}_{c,s}T_i)$, etc.

Across each flow variation the corresponding quantity and/or time variations are monitored and classified. There is evident, that a basic state across particular flow variation should reflect either no variation of the inspected scalar variable, or just a tolerable one captured between some acceptable bounds. Being judged by company management, such state is mostly desired and expected. It is called *in-control state*. In principle, we may consider the *in-control state* either by specifying suitable events A_n or by setting corresponding bounds, thresholds respectively, which may be un-symmetric to neutral state, and which are directly applicable during processing the problem-oriented data. Except one in-control state, it is necessary to define other states by setting their bounds. These bounds depend usually upon severity of flow variations from managerial point of view. Such other states are called *out-of-control states*.

For a generic scalar variable representing any flow variation, p_1 gives its in-control state probability, s gives its total number of states considered, and p_i, i=2,..., s denote probabilities of all its out-of-control states. Following the basic assumption that these s states form a set of complete events, it holds

$$\sum_{i=1}^{s} p_i = 1, \text{ hence } \sum_{i=2}^{s} p_i = 1 - p_1. \tag{6}$$

Now using (4), we may express entropy of such flow variation in the following form

$$h(p_1,..., p_s) = - p_1 \log_2(p_1) - \sum_{i=2}^{s} p_i \log_2(p_i). \tag{7}$$

Following [5], we will consider a general supplier-customer system with its interface included, which consists of set of products $\{P_1, ..., P_n\}$, each P_i having r_i flow variations, i=1,..., n, and each flow variation being represented by s_{r_i} mutually disjunctive states, i.e. one desired in-control state and the others out-of-control states. On the base of (9), thus we may express the entropy of supplier-customer system as follows

$$H = -\sum_{i=1}^{n} \sum_{j=1}^{r_i} (p_{ij1}\log_2(p_{ij1}) - \sum_{k=2}^{s_{r_j}} p_{ijk}\log_2(p_{ijk})), \tag{8}$$

where p_{ij1} stands for probability of in-control state belonging to the j-th flow variation from r_i ones being considered at product P_i. Whereas p_{ijk} are corresponding all out-of-control states introduced.

However, there are two ways or approaches how to tackle out-of-control states. The first one follows processing H with respect of conditional probabilities expressing occurrence of out-of-control states. The second one seems computationally simpler since it is realizable by proper filtering of problem-oriented data.

In the first approach, it is useful to express out-of-control states and their probabilities p_{ijk} as conditional ones, in particular conditioned by their compound event to the corresponding in-control state. Thus, on the base of (8), and using property that all states introduced are mutual disjunctive ones, we may write

$$\sum_{k=2}^{s_{r_j}} p_{ijk} = 1 - p_{ij1}, \text{ or equivalently } (1 - p_{ij1})^{-1} \sum_{k=2}^{s_{r_j}} p_{ijk} = 1, \tag{9}$$

which yields

$$p_{ijk} = (1 - p_{ij1})\, q_{ijk}\,, \quad \text{and} \quad \sum_{k=2}^{s_{rj}} q_{ijk} = 1, \tag{10}$$

where q_{ijk} are conditional probabilities introduced.

Substituting (11) into (10) gives

$$H = -\sum_{i=1}^{n} \sum_{j=1}^{r_j} (p_{ij1}\log_2(p_{ij1}) - \sum_{k=2}^{s_{rj}} (1 - p_{ij1}) q_{ijk} \log_2((1 - p_{ij1})q_{ijk})), \tag{11}$$

and finally, we obtain

$$H = -\sum_{i=1}^{n} \sum_{j=1}^{r_j} (p_{ij1} \log_2(p_{ij1}) - (1 - p_{ij1})\log_2(1 - p_{ij1}) - (1 - p_{ij1}) \sum_{k=2}^{s_{rj}} q_{ijk} \log_2(q_{ijk})). \tag{12}$$

The operational complexity of supplier-customer system measured by entropy (12) depends upon the set of mutually disjunctive states introduced with all flow variations considered and their corresponding probabilities. Inspecting its structure we notice that it can be divided into three additive terms H_1, H_2 and H_3, respectively

$$\begin{gathered} H = H_1 + H_2 + H_3, \\ H_1 = -\sum_{i=1}^{n} \sum_{j=1}^{r_j} p_{ij1} \log_2(p_{ij1}), \quad H_2 = -\sum_{i=1}^{n} \sum_{j=1}^{r_j} (1 - p_{ij1})\log_2((1 - p_{ij1}), \\ H_3 = -\sum_{i=1}^{n} \sum_{j=1}^{r_j} (1 - p_{ij1}) \sum_{k=2}^{s_{rj}} q_{ijk} \log_2(q_{ijk}), \end{gathered} \tag{13}$$

where H_1 represents the amount of information required to describe the system is in desired in-control states, H_2 represents the amount of information required to describe the system is out of desired in-control states, and H_3 represents the additional amount of information required to describe the system is in all out-of-control states considered.

There is evident that these quantities provide more useful information than H only. At first, the ratio H_2/H_1 describes information-theoretic fraction showing how much is the system out of desired in-control states related to being in all desired in-control states. At second, the quantity H_3 could be tackled as stronger information-theoretic measure of operational complexity of the supplier-customer system, because it gives an expected amount of information to describe the extent to which the system occurs in all out-of-control states considered.

In the second approach, we simply use formula (8) just with adopted set of events, which depends upon defined thresholds. Since it is realizable directly during processing of problem-oriented database we prefer this approach. As to the computer implementation, we have developed a coupled system of programs in Java and Mathematica, Wolfram Research, Inc. The processing of problem-oriented database is realized by Java programs. On the other hand, the computation of empirical distribution of events and evaluation of entropy of the system is implemented by Mathematica notebook.

4 Case study – variable threshold period

As a case study for investigation of variable threshold period influence upon evaluation of entropy measured operational complexity of supplier-customer system, we have selected the company Lubricant Ltd. (s.r.o.) being SME ranged one. We are focused upon perturbations between terms of planned and actual deliveries of goods over the whole year 2011. The company has a dominant supplier whose behavior causes problems notwithstanding the settled contracts. So, we have considered primarily perturbations within the quantity $T_p - T_s$. The problem-oriented database consists of 452 data. It was pre-processed by Java program `EnComP1mma.java`, first. And further, the corresponding data are fetched into `EnComP2mma.nb` Mathematica notebook for numerical calculations and creation of graphical outputs.

Figure 1 show the corresponding source data, i.e. $T_p - T_s$ differences, in particular. Figure 2 displays its empirical distribution function. Inspecting the data we may simply observe that the perturbations are significantly clustered at 7-day long periods, i.e. supply delivery delays show pre-dominant week-long anomalies. There was

just an impetus for further investigation of an effect of variable threshold periods upon entropy evaluation of this supplier-customer relation. We have defined an in-control-state period $[b_d, b_u]$ by setting thresholds b_d, b_u being lower and upper thresholds for quantity $T_p - T_s$, respectively. As to Figure 1 and 2, they have $[b_d, b_u] = [0, 0]$.

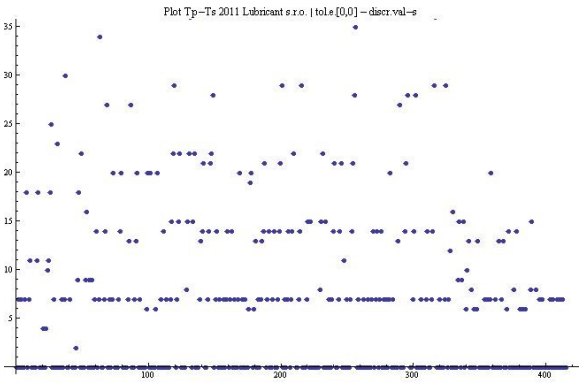


Figure 1 $T_p - T_s$ primer data set

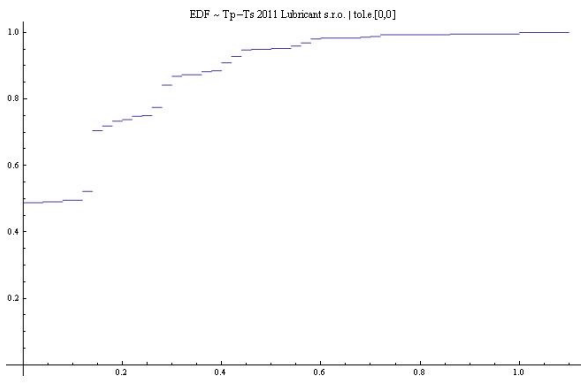


Figure 2 Empirical distribution function of $T_p - T_s$

Next, we shall apply the threshold period $[b_d , b_u] = [0, 7]$ in days. The following results are displayed on Figure 2 and 3, giving the filtered $T_p - T_s$ data set by the given threshold period, and corresponding empirical distribution function, too.

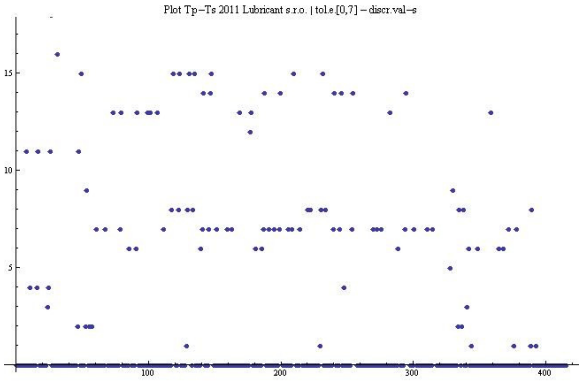


Figure 3 $T_p - T_s$ filtered by $[b_d, b_u] = [0,7]$

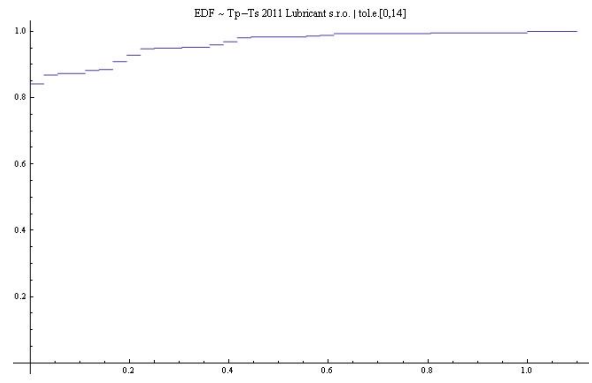


Figure 4 EDF of $T_p - T_s$ filtered by $[b_d, b_u] = [0, 7]$

Further, we have applied the threshold period $[b_d, b_u] = [0, 14]$ in days. The following results are displayed on Figure 4 and 5, giving the similar data, i.e. filtered $T_p - T_s$ data set by the given threshold period, and corresponding empirical distribution function, too.

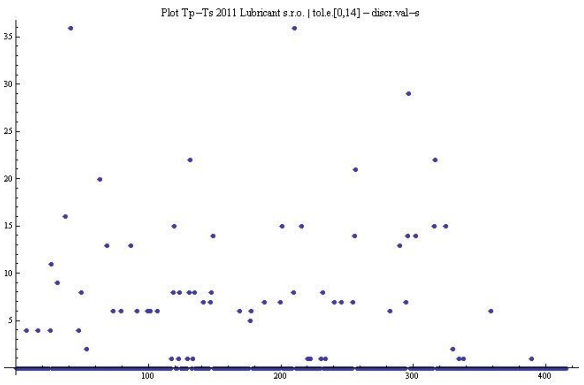


Figure 5 $T_p - T_s$ filtered by $[b_d, b_u] = [0, 14]$

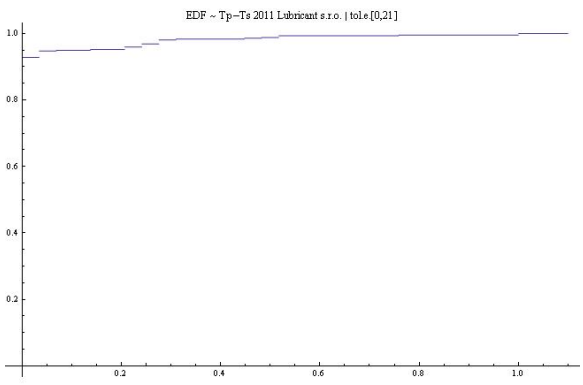


Figure 6 EDF of $T_p - T_s$ filtered by $[b_d, b_u] = [0, 14]$

Finally, we have applied the threshold period $[b_d, b_u] = [0, 21]$ in days, and we have got similar pictures. Since limited number of paper pages we do not present them here, but we present the picture which shows dependence of calculated entropy ratio, i.e. H/I_u evaluated by formulas (5) and (8), upon threshold period of in-control state bounds $[b_d, b_u] = [0, b]$, $b = 0,1,..,14$ days. The corresponding result is depicted on Figure 7. The graph shows strong dependence of relative entropy, i.e. entropy ratio $H(b)/I_u$, upon in-control state threshold period $[0, b]$ being accepted, where $H(b)$ denotes entropy calculated by (8) from pre-processed data taking into account the given thresholds, i.e. 0 and b. Such results can be directly used by management of company for negotiation and settlement proper details of supplier-consumer contracts.

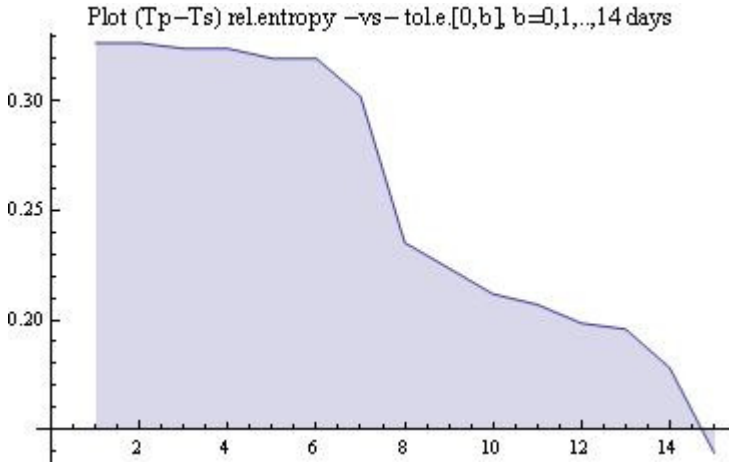


Figure 7 Entropy ratio H/I_u of $T_p - T_s$ filtered by $[0, b]$, $b = 0,1,..,$ 14 days

5 Conclusion

The measure of operational complexity based upon entropy provides versatile instrument for supplier-customer system analysis. In case studies presented, we applied unified approach for analysis of lead time variations of product-line main products at four SME-ranked firms. As a new feature we investigated the role of varying threshold periods with un-symmetric bounds. The results are briefly discussed, and possibilities of their direct application for managerial decision making are given.

References

[1] Lau, A.H.L., Lau H-S.: Some two-echelon supply-chain games improving from deterministic-symmetric-information to stochastic-unsymmetric-information. *EJOR* **161** (2005), 203-223.

[2] Lukáš, L.: *Probabilistic models in management Inventory theory and statistical description of demand* (*Pravděpodobnostní modely v managementu Teorie zásob a statistický popis poptávky*, in Czech). Publ. Academia, Prague, 2012, ISBN 973-80-200-2005-5, pp.208.

[3] Lukáš, L., Hofman, J.: Application of entropy for measuring supplier-customer system complexity – a quantitative tool for firm financial management (Použití entropie k měření složitosti dodavatelsko-odběratelského systému - kvantitativní nástroj využitelný finančním managementem podniku, in Czech). In: Hrdý, M. et al.: *Complex solution of theoretical and applied problems of financing SME in EU market environment* (*Komplexní řešení teoretických a aplikačních problémů financování malých a středních podniků v podmínkách tržního prostředí Evropské unie*, in Czech). Vyd.ZČU, Plzeň, 2008, ISBN 978-80-7043-746-9, pp.172, Chap.7, 73-82.

[4] Prochorov, Ju. B., Rozanov, Ju., A.: *Theory of probability* (*Teorija verojatnostej*, in Russian). Nauka, Moskva, 1967.

[5] Sivadasan, S., Efstathiou J., Calinescu A., Huaccho Huatuco L.: Advances on measuring the operational complexity of supplier-customer systems. *EJOR* **171** (2006), 208-226.

[6] Soldanová, K.: *Measuring of supplier-customer relations complexity using quantitative measures in Lubricant, Ltd*. (*Měření složitosti dodavatelsko-odběratelských vztahů pomocí kvantitativních měr ve firmě Lubricant, s.r.o*, in Czech). Thesis, Univ. of West Bohemia in Pilsen, ZČU/FEK, Pilsen, 2012.

Collusive general equilibrium between aggregated industries

Milan Horniaček[1]

Abstract. We analyze an infinite horizon difference game between four aggregated industries - production of producer goods, production of consumption goods, federation of labor unions, and commercial banking sector. A strict strong perfect general equilibrium is the applied solution concept for the game. It requires that no coalition in no subgame can weakly Pareto improve the vector of payoffs of its members. We formulate and prove the sufficient condition for its existence.

Keywords: collusion, difference game, general equilibrium. strict strong perfect general equilibrium.

JEL Classification: C73, D43, D51.

AMS Classification: 91A-06.

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,[2] or [1] for description of these models) follow the paradigm of perfect competition. The same holds also for computable general equilibrium models focusing on typical oligopolistic industries (see [4] for a recent example). Thus, general equilibrium models that incorporate the interaction between imperfectly competitive firms are needed. So far, the paper by Dierker and Grodal [3] is the most comprehensive result in this direction. In their static model, they pay a great attention to the issue of the objective of an imperfectly competitive firm.

In the present paper, we construct an infinite horizon aggregated general equilibrium model with imperfectly competitive producers. The model is aggregated because industries play the role of firms, banks, and labor unions. There are four aggregated industries: production of producer goods, production of consumption goods, federation of labor unions, and commercial banking sector.

Consumers do not behave strategically. They make their decisions on the basis of maximization of average discounted utility. Therefore, we do not include them in the set of players in the game.

A strict strong perfect general equilibrium (henceforth, SSPGE) is the equilibrium concept applied to the model. It requires that there does not exist a coalition of players that can weakly Pareto improve the vector of continuation payoffs of its members in some subgame by a coordinated deviation. It is a refinement of Rubinstein's [5] concept of a strong perfect equilibrium. We formulate and prove a sufficient condition of its existence in our model.

[1]Comenius University in Bratislava, Faculty of Social and Economic Sciences, Institute of Public Policy and Economics, Mlynské luhy 4, 821 05 Bratislava, e-mail: milan.horniacek@fses.uniba.sk

2 Model

Throughout the paper, $\mathbb{N}$ ($\mathbb{R}$) denotes the set of positive integers (real numbers). We endow each finite dimensional real vector space with the Euclidean topology and each infinite dimensional Cartesian product of finite dimensional real vector spaces with the product topology.

We denote the analyzed difference game with discount factor $\delta \in (0,1)$ (that is common for all players) by $\Gamma(\delta)$. It is played in periods numbered by positive integers. There are four players: the industry producing producer goods - player J, the industry producing consumption goods - player C, the federation of labor unions - player L, and the commercial banking sector - player B. $G = G_J \cup G_C \cup G_L \cup G_B$ is the finite set of goods in the model. G_J (G_C, G_L, G_B) is the set of producer goods (consumption goods, labor services, banking products). In order to simplify the exposition of the model, we assume only one type of deposit account and only one type of credit for consumers. For each $g \in G$, $y_g^{\max} > 0$ is the maximal amount of good g that can be produced (provided) in a single period. $G_J = G_{J+} \cup G_{J-}$ and $G_C = G_{C+} \cup G_{C-}$, where G_{J+} (G_{C+}) is the set of capital (durable consumption) goods and G_{J-} (G_{C-}) is the set of non-durable producer (consumer) goods. For each $g \in G$, $T_g \in \mathbb{N}$ is the durability of good g. We have $T_g > 1$ for each $g \in G_{J+} \cup G_{C+}$ and $T_g = 1$ for each $g \in G_{J-} \cup G_{C-} \cup G_B$. For each $g \in G$, $p_g^{\max} > 0$ is the upper bound on its price in any period.

I is the finite set of infinitely living consumers. For each $i \in I$ and each $t \in \mathbb{N}$, $\omega_i(t) = (\omega_{ig}(t), g \in G_L) \in \prod_{g \in G_L} \left[0, y_g^{\max}\right]$ is i's endowment by labor services in period t. $I_L \subseteq I$ is the non-empty set of members of L. For each of them L concludes labor contracts with J, C, and B on their behalf. L also pays unemployment benefits to its unemployed members. In order to do so, it collects part $\beta \in (0,1)$ from wages of all employed members to the unemployment fund. We assume that J, C, and B are aggregates of limited liability companies whose shares are not traded. (Despite this we use the term "dividends" for payments that they make to their owners.) Consumers do not behave strategically. Taking employment contracts concluded on their behalf by L as given, they make their decisions on consumption, savings, and borrowing from B on the basis of maximization of their average discounted utility. Therefore, we do not include consumers in the set of players in $\Gamma(\delta)$. For each consumer, his average discounted utility is computed using his continuous single period utility functions (which is the same for each period) that represents his locally non-satiated and strictly convex preferences. There is non-zero vector $x^{\min} \in \prod_{g \in G_{C-}} \left[0, y_g^{\max}\right)$ that enables each consumer to survive.

For each $k \in \{J, C, B\}$, k's technology correspondence Y_k assigns to each input vector

$$\begin{aligned} z_k &= (((z_{kg\tau}, \tau \in \{1, ..., T_g\}), g \in G_{J+}), (z_{kg}, g \in G_{J-} \cup G_L \cup G_B)) \\ &\in \prod_{g \in G_{J+}} \left[0, y_g^{\max}\right]^{T_g} \times \prod_{g \in G_{J-} \cup G_L \cup G_B} \left[0, y_g^{\max}\right] \end{aligned}$$

the set $Y_g(z_g)$ of output vectors that it can produce from z_k. For each input vector z_k, $Y_k(z_k) \subseteq \prod_{g \in G_J} \left[0, y_g^{\max}\right]$ if $k = J$, $Y_k(z_k) \subseteq \prod_{g \in G_C} \left[0, y_g^{\max}\right]$ if $k = C$, $Y_k(z_k) \subseteq \prod_{g \in G_B} \left[0, y_g^{\max}\right]$ if $k \in B$.

Assumption 1. *For each $k \in \{J, C, B\}$, Y_k is (i) nonempty-valued ($0 \in Y_k(z_k)$ for each z_k from its domain), (ii) compact-valued, and there exist the non-empty compact $Z_k \subset \prod_{g \in G_{J+}} \left[0, y_g^{\max}\right]^{T_g} \times \prod_{g \in G_{J-} \cup G_L \cup G_B} \left[0, y_g^{\max}\right]$ and $\eta_k > 0$ such that (iii) $z_k \in Z_k$ only if $z_{kg} \geq \eta_k$ for at least one $g \in G_{J+}$, for at least one $g \in G_{J-}$, for at least one $g \in G_L$, and for $g \in G_B$ that is k's deposit account, (iv) $Y_k(z_k) \supset \{0\}$ if and only if $z_k \in Z_k$, and (v) Y_k is continuous on Z_k.*

For each $k \in \{J, C, B\}$ and each $t \in \mathbb{N}$, $z_k(t)$ is the vector of k's inputs in period t resulting from purchases in period $t-1$; $z_k(1)$ is given.

We also assume that C cannot make a positive single period profit when either the wage for each labor service or the price for each non-durable producer good is on its upper bound.

Assumption 2. *(i) For each* $k \in \{J, C, B\}$, $z_k(1) \in Z_k$. *(ii) For each* $t \in \mathbb{N}$, *if* $z_k(t) \in Z_k$ *for each* $k \in \{J, C, B\}$, *then there exist* $y_k \in Y_k(z_k(t))$, $k \in \{J, C, B\}$, and $z_k(t+1) \in Z_k$, $k \in \{J, C, B\}$, *such that the redistribution of* y_k, $k \in \{J, C, B\}$, *and* $\sum_{i \in I} \omega_i(t+1)$ *makes the vectors of inputs* $z_k(t+1)$, $k \in \{J, C, B\}$, *feasible.*

Contracts for delivery of producer goods by J to C or B (with delivery by J to J possible), contracts for delivery of labor services by L to J or C or B, and contracts for granting of credit by B to J or C specify quantity and price. Contracts between B on the one hand and J or C or L on the other hand on deposit conditions specify only the interest rate. A buyer and a seller make simultaneously contract proposals and a contract is concluded if and only if their proposals match. We formally allow contract proposals and contracts with the zero traded quantity. All payments made to J, C, and the consumers are deposited into their deposit accounts.

Each period is divided into five phases. In the first phase, J makes a proposal of contract for delivery of producer goods to C and B and C and B make a proposal of contract for purchase of producer goods to J. C makes a decision on production of consumption goods. These decisions should be compatible with available inputs. Also, B makes the proposal of a contract on deposit conditions to J, C, and L for the current period and it announces the deposit interest rate for consumers' deposit accounts and the interest rate for consumer credit for the current period. Similarly, J, C, and L propose a contract on deposit conditions for the current period to B. In the second phase, production and delivery of producer goods by J according to concluded contracts and production of consumer goods and announcement of their prices by C takes place. In the third phase, sale of consumption goods takes place. In the fourth phase, wages are paid and L collects contributions to the unemployment fund, and pays unemployment benefit to each of its unemployed members. Each unemployed member receives the sum equal to the sum collected from the employed members divided by the number of unemployed members. In the fourth phase, B makes a proposal of contract for granted credit to J and C.J and C make a proposal of contract for received credit to B. In the fifth phase, the credits granted in the preceding period are repaid (including the interest). If J or C is unable to repay its debt, B appropriates its profits in the following periods until the debt is repaid. (For discount factor close enough to one, this appropriation rule does not lead to moral hazard on the part of J or C.) If a consumer is unable to repay his debt, B appropriates his income minus the sum needed to buy $x^{\min}$ minus the contribution to unemployment fund (if the resulting sum is negative, nothing is appropriated) in each following period until the debt is repaid. In the fifth phase, J, C, and B choose the dividend per share (their sum should not exceed the profit after deduction of the appropriated sum) and pay them. Also, L makes a proposal of contract for delivery of labor services (in the following period) to J, C, and B and J, C, and B make a proposal for purchase of labor services to L. J, C, and B should have (in the case of J and C, in its deposit account) the sum sufficient for paying wages for the labor services it intends to contract.

All players observe all past actions of all players. Thus, $\Gamma(\delta)$ is a game with perfect information of players.

It follows from parts (iii) and (iv) of Assumption 1 that if $z_J(t) \notin Z_J$ or $z_B(t) \notin Z_B$, then, starting from period $t+1$, no economic activity takes place. (In our model we do not consider foreign help or other ways how to revive the economy.) In addition, we assume that no economic activity takes place since period $t+1$ also when $z_C(t) \notin Z_C$. This can be interpreted as the inability of workforce to survive.

H is the set of non-terminal (i.e., finite) histories in $\Gamma(\delta)$. We restrict attention to pure strategies in $\Gamma(\delta)$. A pure strategy of player $k \in \{J,C,B,L\}$ assigns its decision to each $h \in H$ after which he makes (according to the above description of phases) a decision. We denote the set of his pure strategies by S_k and let $S = \prod_{k\in\{J,C,B,L\}} S_k$and $S_D = \prod_{k\in D} S_k$ for each $D \in 2^{\{J,C,B,L\}} \setminus \{\varnothing\}$. Since all players' choices are from compact sets, using Tychonoff's theorem, S is compact.

In the computation of consumer price index (henceforth, CPI) in our model we use prices of all goods in G_C, with their quantities sold in the first period as weights. Real values are nominal values divided by CPI. For each $k \in \{J,C,B,L\}$, $\pi_k : S \to \mathbb{R}$ is k's payoff function in $\Gamma(\delta)$. (Although its functional values depend on δ, we do not include δ explicitly among its arguments.) For $k \in \{J,C,B\}$, $\pi_k(s)$ is the average discounted sum of real dividends paid by k when the players follow s; $\pi_L(s)$ is the average discounted sum of real wages when the players follow s.

We denote by $S_=$ the subset of S with the following property: $s \in S$ belongs to $S_=$ if and only if it prescribes (both after non-terminal histories consistent with it and after non-terminal histories that are not consistent with it) only contract proposals that lead to conclusion of a contract. Clearly, for each $s \in S\setminus S_=$ there exists $s' \in S_=$ that generates the same outputs, prices, traded quantities, wages, dividends, and payoffs.

For each $h \in H$, $\Gamma_{(h)}(\delta)$ is the subgame of $\Gamma(\delta)$ following history h. For any set or function defined for $\Gamma(\delta)$ its restriction to $\Gamma_{(h)}(\delta)$ is indicated by the subscript (h).

We denote by $S_{=+}$ the subset of $S_=$ whose elements generate input vectors belonging to Z_k, $k \in \{J,C,B\}$, in each subgame at the beginning of which input vectors belong to Z_k, $k \in \{J,C,B\}$. Note that $S_{=+}$ is nonempty and compact and players' payoff functions are continuous on it.

Assumption 3. *There exists $\overline{\varepsilon} > 0$ such that for each $\varepsilon \in (0,\overline{\varepsilon}]$, there is $s \in S$ such that in each $t \in \mathbb{N}$, after any $h \in H$ leading to the first phase of t and generating $z_k \in Z_k$ for each $k \in \{J,C,B\}$ at the beginning of t, each player has single period payoff no lower than ε.*

We end this section with the definition of an SSPGE.

Definition 1 *Strategy profile $s^* \in S$ is an SSPGE of $\Gamma(\delta)$ if (i) there do not exists $h \in H$, $D \in 2^{\{J,C,B,L\}} \setminus \{\varnothing\}$, and $s_D \in S_{D(h)}$ such that*

$$\pi_{k(h)}\left(\left(s_D, s^*_{-D(h)}\right)\right) \geq \pi_{k(h)}\left(s^*_{(h)}\right)$$

for each $k \in D$ with strict inequality for at least one $k \in D$ and (ii) there do not exist $h \in H$ and $s \in S_{(h)}$ such that $\pi_{k(h)}(s) \geq \pi_{k(h)}\left(s^*_{(h)}\right)$ for each $k \in \{J,C,B,L\}$ with strict inequality for at least one $k \in \{J,C,B,L\}$.

3 Existence of an SSPGE

For each $\delta \in (0,1)$ and $\varepsilon \in (0,\overline{\varepsilon}]$, let $S(\delta,\varepsilon)$ be the subset of $S_=$ with the following property: $s \in S(\delta,\varepsilon)$ if and only if (i) in each $\Gamma_{(h)}(\delta)$ (including each subgame following $h \in H$ that is not consistent with s) starting in the first phase of a period and with $z_k \in Z_k$ for each $k \in \{J,C,B\}$ at the beginning of $\Gamma_{(h)}(\delta)$, it generates a strictly Pareto efficient vector of average discounted payoffs in $\Gamma_{(h)}(\delta)$ and (ii) in each $t \in \mathbb{N}$, after any $h \in H$ leading to the first phase of t and generating $z_k \in Z_k$ for each $k \in \{J,C,B\}$ at the beginning of t, each player has single period payoff no lower than ε. Recall that $S_{=+}$ is nonempty and compact and players' payoff functions are continuous on it. Thus, taking into account Assumption 3, for each $\varepsilon \in (0,\overline{\varepsilon}]$ there exists $\delta'(\varepsilon)$ such that $S(\delta,\varepsilon) \neq \varnothing$ for each $\delta \in \left[\delta'(\varepsilon), 1\right)$.

Proposition 1 *For each* $\epsilon \in (0, \overline{\varepsilon}]$ *there exists* $\underline{\delta}(\varepsilon) \in (0,1)$ *such that for each* $\delta \in [\underline{\delta}(\varepsilon), 1)$ *and each* $s \in S(\delta, \varepsilon)$ *there is* $s^* \in S_=$ *that is an SSPGE of* $\Gamma(\delta)$ *and generates the same terminal history as* s.

Proof. (With respect to space limitations, we only outline the proof.) We set $\lambda \in (0,1)$ such that a deviation by any coalition other than the grand one leading to history h is punished by $s^*_{(h)}$ that differs from $s_{(h)}$ in two ways: for T periods the single period payoff of at least one deviator is no greater than $\lambda\varepsilon$ and then it equals ε. In the remainder of the proof, we call the first T periods of the punishment of deviator $k \in \{J, C, B, L\}$ k's punishment phase. A new deviation by a coalition other than the grand one when the punishment phase of all previously punished players passed is treated in the same way as a deviation after the history that contains no deviation. During his punishment phase, k is punished by each player who is not punished himself. The punishment is carried out by increased input prices and/or decreased output prices, without changing outputs and traded quantities. A new deviation does not interrupt the punishment of the previous deviation but a new deviation is also punished. (Thus, the above described punishment are sufficient for punishing a deviation by a coalition other than the grand one whose members start their deviations in different time periods.) We choose T large enough to ensure that for δ close enough to one the above described punishments indeed make deviations by all coalitions other than the grand one unprofitable. (This is the first factor determining $\underline{\delta}(\varepsilon)$.) Deviations during the punishment phase of some player in period t prevent J or C from producing or B from providing banking services since period $t+1$. Thus, they lead to zero single period payoffs of all players since period $t+1$ or $t+2$. Therefore, they are unprofitable for δ close enough to one. (This is the second factor determining $\underline{\delta}(\varepsilon)$.) By construction, $s^*_{(h)}$ gives the strictly Pareto efficient vector of average discounted payoffs in $\Gamma_{(h)}(\delta)$ for each $h \in H$. Hence, the grand coalition cannot weakly Pareto improve the vector of average discounted payoffs by a deviation. ■

4 Conclusions

In the present paper, we have formulated the sufficient condition for the existence of an SSPGE - a general equilibrium between aggregated industries, which is immune to deviations by all coalitions in all subgames. The payoff of each production industry and the commercial banking sector is equal to the average discounted sum of real dividends of its owners. The payoff of the federation of labor unions is equal to the average discounted sum of real wages. By definition, in each subgame, the equilibrium payoff vector in an SSPGE is strictly Pareto efficient. This implies that it is not possible to increase average discounted real income of any consumer without decreasing average discounted real income of another consumer.

Acknowledgement 1 *The research reported in the present paper was supported by the grant VEGA 1/0181/12 from the Slovak Ministry of Education, Science, Research, and Sport.*

References

[1] Arrow, K.J., and Hahn, F.H.: *General competitive analysis*. Holden and Day, San Francisco, 1971.

[2] Debreu, G: *Theory of value.* Yale University Press, New Haven, 1959.

[3] Dierker, E., and Grodal, B.: The price normalization problem in imperfect competition and the objective of the firm. *Economic Theory* **14** (1999), 257-284.

[4] Miťková, V.: Modeling the crisis in automobile industry - the applied CGE study. In: *Proceedings of the Mathematical Methods in Economics 2009* (Brožová, H., and Kvasnička, R., eds.). Czech University of Life Sciences, Faculty of Economics and Management, Prague, 2009, 221-226.

[5] Rubinstein, A.: Strong perfect equilibrium in supergames. *International Journal of Game Theory* **9** (1980), 1-12.

Convexity in stochastic programming model with indicators of ecological stability

Michal Houda[1]

Abstract. We develop an optimization model dealing with construction expenses that are prescribed as a result of the EIA (Environmental Impact Assessment) process. The process is an obligatory part of every large construction project and evaluates possible influences of the project to the environment, including population health, natural and other socio-economic aspects; the result of the process is a set of recommendation and arrangements the construction must meet.

Our optimization model incorporates uncertainties in model parameters; we represent them through their probabilistic distribution. Furthermore, to overcome a problem with quantifying subjective utility function of ecological impacts, we measure them by so-called indicators of ecological stability. The resulting problem is stochastic programming problem formulated as (C)VaR model used traditionally in finance area. In our contribution we deal with convexity properties of this problem – these are especially important from the theoretical as well as from the computational point of view. We propose a series of assumptions to the problem that ensure convexity of the final set of feasible solutions and/or the objective function.

Keywords: EIA process, indicator of ecological stability, stochastic programming, value-at-risk models, convexity.

JEL classification: C44
AMS classification: 90C15

1 Introduction

In our previous contributions [6, 7] we started to analyze optimization models that deal with expenses connected with large, especially line constructions. We will continue in this direction and deal with convexity issues that are indispensably important from the theoretical and practical point of view.

1.1 Problem formulation

We recall here shortly the subject matter we deal with. During several last decades we notice a very considerable growth of number of small as well as big engineering constructions. This growth is accompanied by deeper regulations of the rules under which the constructions are rising. One of the most important changes compared to the past is the growing emphasis to the environment.

Every important construction must obey the so-called Environmental Impact Assessment (EIA) imposed by European Union Law (see [3]). This process is in fact evaluation of impacts of the construction to the environment and human healthy, divided into two classes and several categories:

- influence of the construction to the *human healthy*: noise pollution, air pollution and social-economic (comfort) factors (life conditions, transport services and loads, emotions);
- influence of the construction to to the *environment*: air, climate, water, land, forests, natural resources, flora, fauna, ecosystems, landscape, systems of ecological stability and also the tangible property or cultural heritage.

[1]Institute of Information Theory, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8, Czech Republic, e-mail: houda@ef.jcu.cz

The detailed list of the categories is given by [4] and the summary is also presented in [6]. The impacts are of very manifold nature – furthermore each of the factors involves many inputs and outputs and their evaluations is not simple. This all contribute to the fact that modelling the optimization problem is not easy and so is for calculation.

In the next section we recall the formulation of the mathematical model [7] and state some of its basic properties. We then focus on the convexity of the set of feasible solutions; we recall some known results from the literature that deals with convexity issues of chance-constrained probabilistic problems that can be more or less applied to our model.

2 Model description

The traditional cost-minimizing optimization model is not considered in this paper because of difficulties with non-linear utility function (representing the environmental impacts of the construction). Instead we consider model where costs are incorporated as the constraint with the right-hand side representing the budget limitation.

2.1 Uncertain optimization model

Denote $x \in X \subset \mathbb{R}^n$ the set of the possible compensations and arrangements ready to be used in some construction. The values of the variable x can be discrete or continuous according to the nature of the arrangement: they can be binary variables (to realize or not the arrangement), discrete (representing possible variants of the arrangements) as well as continuous (dimensions and other measures of the arrangement).

Let $\xi \in \Xi$ be the random vector representing uncertainty factors, with $\Xi \subset \mathbb{R}^s$ being the predefined support of ξ. The uncertainty factors represents traffic intensities (when building transport constructions), efficiency of compensating constructions, quantification of subjective criteria, accidents, etc. From the computational point of view, the probability distribution of ξ has to be known in advance.

The actual expenses of all arrangements are represented by the cost function $c : X \times \Xi \to \mathbb{R} : (x; \xi) \mapsto c(x; \xi)$. Suppose c to be a linear function of x, i. e., $c(x; \xi) = c^T x$ where $c \in \mathbb{R}^n$ are constant unit costs of the arrangements. This simplification is not crucial in our setting; possible dependence on random factor can be (later) incorporated in the probabilistic part of the model. Let furthermore B be the budget limit on the expenses.

The utility function $u : X \times \Xi : (x; \xi) \mapsto u(x; \xi)$ represents the factors of subjective and evaluative character. The quantification of this function is generally difficult. In our paper, we use the approach in which we replace the subjective utility function by a function based on the indicators of ecological stability.

We are now ready to write down the uncertain formulation of our model

$$\text{maximize } u(x; \xi) \text{ subject to } c^T x \leq B, x \in X_0, \tag{1}$$

where $X_0 \subset \mathbb{R}^n$ represents all deterministic constraints of the model (except costs).

2.2 Chance-constrained optimization model

We first deal with subjective utility function: as already outlined we replace it by indicators of ecological stability [4]. More precisely, let $i : \mathbb{R}^n \times \Xi \to \mathbb{R}^I : (x; \xi) \mapsto i(x; \xi)$ be a function representing the values of the EEA indicators, representing the functional dependence of the indicator value on decision vector x and random factor ξ. Next, we introduce a parameter L representing a required limit on values of i, and weights $w \in [0; 1]^I$ representing relative importance of each of indicators. The probabilistic information about ξ is incorporated model through the probabilistic constraint; we write the resulting final model in the form of

$$\text{maximize } L \text{ subject to } \mathbb{P}\{w^T i(x; \xi) \geq L\} \geq 1 - \varepsilon,\ c^T x \leq B, x \in X_0, \tag{2}$$

where $1 - \varepsilon$ ($\varepsilon \in [0; 1]$) represents some prescribed probability of fulfilling the required limit on indicators. The limit L is considered joint here and compensations for indicator values are allowed. Other approaches,

such as treating the limits individually (or by groups of indicators), are possible but imply more unpleasant issues concerning convexity of the feasibility set. We will treat convexity of the set of feasible solutions in the next section.

3 Convexity analysis

3.1 Notion of generalized convexity

To analyze convexity properties of probability constraint of (2) we recall some results dealing with general forms of probabilistic constraints. Denote

$$X_\varepsilon := \left\{x \in X : \mathbb{P}\{g(x;\xi) \geq 0\} \geq 1-\varepsilon\right\} \tag{3}$$

the set with probabilistic constraint where $g : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^d$ is vector-valued mapping. If g is convex then the sets X_1, X_0 are convex. Unfortunately these sets are not of great interest: they represent existence and almost sure constraint fulfilling. The main "classical" result belongs to Prékopa [10] developed later especially by Borell [1] and [2]. To state the result we need introduce the notion of r-concave functions and measures.

Definition 1. A function $f\colon \mathbb{R}^d \to (0;+\infty)$ is called *r-concave* for some $r \in [-\infty;+\infty]$ if

$$f(\lambda x + (1-\lambda)y) \geq [\lambda f^r(x) + (1-\lambda)f^r(y)]^{1/r} \tag{4}$$

is valid for each $x, y \in \mathbb{R}^d$ and each $\lambda \in [0;1]$. The cases $r = -\infty, 0, +\infty$ are treated by continuity.

Among the r-concave functions we specify some important special cases which we enumerate in the following list:

- $r = -\infty$: the right hand side (RHS) of (4) is equal to $\min\{f(x), f(y)\}$ and f is called *quasi-concave* in this case. Quasi-concave functions play very important role in the context of probabilistic programming as we will see later;
- $r \in (-\infty; 0)$: f is r-concave function with negative r if f^r is convex function
- $r = 0$: RHS of (4) is equal to $f^\lambda(x)f^{1-\lambda}(y)$ and the function f is called logarithmically concave or log-concave (as $\log f$ is concave function). Log-concavity is useful property in stcohastic programming; many prominent probability densities share this property and the original Prékopa's results are formulated for log-concave functions (even that proofs are more general);
- $r = 1$: corresponds to classical notion of concavity;
- $r \in (0;+\infty)$: f is r-concave function with positive r if f^r is concave function;
- $r = +\infty$: the right hand side (RHS) of (4) is equal to $\max\{f(x), f(y)\}$ and f is called *quasi-convex* in this case.

If f is r^*-concave for some r^* then it is r-concave for all $r \leq r^*$. In particular, every r-concave function is also quasi-concave.

Definition 2. Probability measure $\mathbb{P}$ (on $\mathcal{B}(\mathbb{R}^s)$) is called *$r$-concave* if for any Borel convex set $A, B \subset \mathcal{B}(\mathbb{R}^s)$ such that $\mathbb{P}(A) > 0, \mathbb{P}(B) > 0$, and any $\lambda \in [0;1]$ we have

$$\mathbb{P}(\lambda A + (1-\lambda)B) \geq [\lambda \mathbb{P}^r(A) + (1-\lambda)\mathbb{P}^r(B)]^{1/r}$$

cases $r = -\infty, 0, +\infty$ are treated by continuity.

It is worth to note three important properties of r-concave probability measures:

- r-concave probability measure has r-concave distribution function;
- non-degenerate quasi-concave measure $\mathbb{P}$ on $\mathbb{R}^s$ with dimension s of the support has a density;

- r-concave density is equivalent to $\frac{r}{1+mr}$-concave measure on convex subset $\Omega \subset \mathbb{R}^s$ of dimension $m > 0$ (so for $r > -\frac{1}{m}$).

Proposition 1 ([9], Theorems 2.5 and 2.11)**.** *If $\mathbb{P}$ is absolutely continuous (with respect to Lebesgue measure), r-concave measure, and the one-dimensional components of g are quasi-concave functions of (x, ξ) then X_ε is convex set.*

The assumption on r-concavity distribution is equivalent to have γ-concave probability density with $\gamma = \frac{r}{1+rs} \geq -\frac{1}{s}$. Many distribution such that multivariate normal, uniform, Wishart, Beta, Dirichlet, Gamma and others have log-concave (or at least quasi-concave) distribution, or at least for some of theirs parameters (see e. g. [9] for details).

The above proposition plays a basement for more specific results. As r-concavity of probabilistic distribution is not a real obstacle, this is not true for quasi-concavity of function f. The main obstacle is that r-concavity is not preserved under addition for $r < 1$ nor under multiplication for $r < 0$. So more specific structures are examined in the literature in order to overcome these obstacles. We concentrate especially on situation corresponding to linear programming problems with random technology matrix; but the available results are still very rare in this case.

3.2 Problems with random technology matrix

The classical results are that of [8] and [13]:

Proposition 2. *If $\xi \in \mathbb{R}^s$ have non-degenerate multivariate normal distribution, b is constant scalar and $\varepsilon \leq \frac{1}{2}$ then the function*

$$G(x) = \mathbb{P}\{\xi^T x \leq b\}$$

is quasi-concave on $D = \{G(x) \geq \frac{1}{2}\}$, thus X_ε is convex.

The next result is due to [11] and generalizes previous theorem to the case of independent normally distributed rows of technology matrices.

Proposition 3. *If a random matrix T has independent normally distributed rows such that their covariance matrices are constant multiples of each other, and $\varepsilon \leq \frac{1}{2}$ then*

$$G(x) = \mathbb{P}\{Tx \leq b\}$$

is quasi-concave on $D = \{G(x) \geq \frac{1}{2}\}$ thus X_ε is convex.

Recently, [12] found equivalent condition on covariance matrices under the condition of uniform quasi-concavity of functions $G_i(x) := \mathbb{P}\{\xi_{i\cdot} \leq b_i\}$. Other interesting results are given by [5].

Proposition 4 ([5])**.** *Suppose T be the random matrix with pairwise independent normally distributed rows $\xi_{i\cdot}$ indexed by i having means μ_i and variance matrices Σ_i. Then X_ε is convex for*

$$\varepsilon < 1 - \Phi(\max\{\sqrt{3}, u^*\})$$

where Φ is standard normal distribution function,

$$u^* = \max_i 4\lambda_{\max}^{(i)} \left[\lambda_{\min}^{(i)}\right]^{-3/2} \|\mu_i\|,$$

and $\lambda_{\max}^{(i)}$, $\lambda_{\min}^{(i)}$ are the largest and the smallest eigenvalue of Σ_i.

3.3 Application to model with EEA indicators

Probabilistic constraint of the model (2) written in terms of (3) corresponds to the setting

$$g((x, L)^T; \xi) = w^T i(x; \xi) - L.$$

Function g is in general non-linear and verifying convexity become hard in that case: according to Proposition 1 one have to check if indicator function $i(x; \xi)$ is (jointly) quasi concave, i. e. to explore the structure of indicator function. If the indicator function depends linearly on x and ξ one can apply results of Section 3.2:

Theorem 5. *Suppose $i(x;\xi) = x^T A\xi$, random vector ξ having joint normal distribution; assume further $\varepsilon \leq \frac{1}{2}$. Then the feasibility set of* (2) *is convex.*

Proof. Function $i(x;\xi)$ is normally distributed and depend linearly on x and ξ. In this case the assumption of Proposition 2 is fulfilled and the assertion of the theorem follows directly. □

Consider now the case where compensations (lack in one indicator can be superseded by surplus in another one) are prohibited. In this case we impose individual limit L_j for each indicator and the whole considered model becomes in fact the model with joint probability constraint:

$$\text{maximize } L \text{ subject to } \mathbb{P}\{i_j(x;\xi) \geq L_j,\ j \in \mathcal{J}\} \geq 1-\varepsilon,\ c^T x \leq B, x \in X_0, \tag{5}$$

where $\mathcal{J}$ is index set for individual indicator functions. Keep the (affine) linear dependence of ξ and x on indicator function. Nevertheless, we cannot use Kataoka's result (Proposition 2) as the model no longer possesses single constraint row. So the covariance structure of the problem enters into consideration: we have to check if assumption of one of Proposition 3 or 4 is fulfilled. The situation become much more worse without normality because no general results on convexity are not known at present time; one has to confide to approximation techniques in order to find (possibly) inner and outer convex approximation of the problem.

4 Conclusion

We have considered probabilistic formulation of a (non-financial) optimization problem which evaluates impacts of the activities (for example transport constructions) on the environment through so-called indicators of ecological activity. The problem is convex if assumption of normal distribution and affine linear dependence of indicators on decision vector and random factor is fulfilled. If this is not the case, no theoretical results are not known and only convex approximation of the original problem has to be considered.

Acknowledgements

This research was supported by the Czech Science Foundation under the project No. P402/10/0956.

References

[1] Borell, C.: Convex set functions in d-space. *Periodica Mathematica Hungarica* **6** (1975), 111–136.

[2] Brascamp, H. J. and Lieb, E. H.: On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equations. *Journal of Functional Analysis* **22** (1976), 366–389.

[3] Council directive 85/337/EEC of 27 June 1985 on the assessment of the effects of certain public and private projects on the environment. *Official Journal of the European Union*, L 175, 05/07/1985:40–48, June 1985.

[4] EEA indicators [online]. European Environment Agency. Available from World Wide Web: <http://www.eea.europa.eu/data-and-maps/indicators/>, May 2012 [cited 10 May 2012].

[5] Henrion, R. and Strugarek, C.: Convexity of chance constraints with independent random variables. *Computational Optimization and Applications* **41**(2) (2008), 263–276.

[6] Houda, M.: Environmental factors in optimization of traffic line construction expenses. In M. Houda and J. Friebelová, editors, *Proceedings of the 24th International Conference Mathematical Methods in Economics 2010*, University of South Bohemia, Česke Budějovice, 2010, 262–267

[7] Houda, M.: Using indicators of ecological stability in stochastic programming. In M. Dlouhý and V. Skočdopolová, editors, *Proceedings of the 29th International Conference Mathematical Methods in Economics 2011*, Professional Publishing, Praha, 2011, 279–283.

[8] Kataoka, S.: A stochastic programming model. *Econometrica* **31**(1–2) (1963), 181–196.

[9] Prékopa, A.: Probabilistic programming. In A. Ruszczynski and A. Shapiro, editors, *Stochastic Programming. Handbooks in Operations Research and Management Science*, volume 10, Elsevier, Amsterdam, 2003, 267–352.

[10] Prékopa, A.: Logarithmic concave measures with applications to stochastic programming. *Acta Scientiarium Mathematicarum (Szeged)* **32** (1971), 301–316.

[11] Prékopa, A.: Programming under probabilistic constraints with a random technology matrix. *Mathematische Operationsforschung und Statistik, Series Optimizations* **5** (1974), 109–116.

[12] Prékopa, A., Yoda, K. and Subasi, M.-M.: Uniform quasi-concavity in probabilistic constrained stochastic programming. *Operations Research Letters* **39**(3) (2011), 188–192.

[13] Van de Panne, C. and Popp, W.: Minimum-cost cattle feed under probabilistic protein constraints. *Management Science* **9** (1963), 405–430.

Definition of relevant market in beer industry: Application of LA-AIDS model

Tomáš Houška[1], Jaroslav Bil[2]

Abstract. The aim of our article is to investigate whether there exists a relevant product market narrower than beer. We use multilevel budget LA-AIDS model to estimate the demand elasticities of the biggest beer brands and selected beer segments on US scanner data. We then perform a critical loss analysis to investigate if the proposed segments constitute separate relevant markets. Results of our paper can be helpful to competition authorities in their assessment of future mergers in the beer industry.

Keywords: competition, beer, demand estimation, LA-AIDS.

JEL classification: D12, C33, L11
AMS classification: 91B42

1 Introduction

Any competition analysis aiming at market power assessment or investigating the intensity of market competition has to begin with the definition of a market. A competition policy uses a concept of relevant market that identifies boundaries of competition between firms.

Paramount to the relevant market definition is a Hypothetical Monopolist Test. According to this test, a relevant market is constituted by a set of products that could be profitably monopolized - i.e. set of products on which a hypothetical monopolist (a firm controlling the supply of given products) could increase its profits by instituting a small, but significant non-transitory price increase (SSNIP).[3] Relevant market definition is crucial in any merger assessment, as it puts the proposed merger into the perspective relative to the size of the whole (relevant) market and the level of competition on that market.

Beer brewing industry, which has increased in its concentration manifold in last couple of decades, is one of the areas, where relevant market definition for the purpose of merger assessment will play an increasingly important role in the forthcoming future. Most market players have recently grown and increased their market share primarily through mergers rather than organic growth.[4] After several last decades filled with frequent mergers, beer brewing industry has become a global market with few leading players.[5]

In nearly all merger decisions concerning beer brewing industry and issued by the European Commission, the relevant product market was considered to be beer. In this article we investigate, whether the relevant market is not in fact narrower.[6] With increasing concentration of the market caused by mergers to come, results of our analysis could be useful in order to predict the post-merger level of competition on the market and to estimate whether the proposed merger significantly impedes the effective competition on the relevant market.

[1]Office for protection of competition, Chief Economist Department, tř. kap. Jaroše 7, Brno 604 55; Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, Brno 602 00, houska.tomas@mail.muni.cz

[2]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, Brno 602 00, jbil@mail.muni.cz

[3]As significant is considered a price increase by 5-10 %. This price increase is used to approximate the level of market power a given company has. The closer substitute other products are, the less likely a hypothetical monopolist is to institute SSNIP. The relevant market thus results in defining a set of products that consumers consider interchangeable.

[4]Examples of recent major mergers between beer producers are Heineken/BBAG (2003), InterBrew/Becks (2001), InterBrew/Brauerglide (2002), Heineken/Scottisch & NewCastle (2008), InBev/AnHeuser Busch (2008)

[5]Since year 2000, 90 % of the world beer production comes from four biggest producers (Anheuser-Busch InBev, South African Breweries (SAB), Heineken, Carlsberg. For more information see e.g. [7]).

[6]European Commission itself has stated on several occasions that the relevant market could be narrower, but has never used narrower market definition in the final decision. For example in Interbrew/Becks merger decision (2001), the EC states that a separate market for premium beer could exist.

We start our analysis at the most detailed level and estimate the own price elasticity of demand for main beer brands and then for six chosen beer segments. Given these estimates, we analyse the size of the relevant market (whether it is a single beer brand, a beer sector or wider definition) by empirically estimating the Hypothetical Monopolist Test. We focus our analysis only on an off-trade market, which is normally treated as separate market from an on-trade market.[7]

2 Model

For the estimation of beer demand we use an Almost Ideal Demand System (AIDS) of Deaton and Muellbauer [4]. We assume a multi-stage budgeting, where in the first (upper) level the overall beer expenditure is allocated between different beer segments, in the second (bottom) level the demand for individual beer brands is estimated. Multi-stage budgeting specification helps us to solve the curse of dimensionality and allows to obtain more efficient estimates both for beer segments and individual beer brands. On the other hand, it implicitly requires compliance with the weak separability of utility function.[8]

We have investigated the following beer segments: dark beers (known for their taste), imported beers (specific by their origin), light beers (with low caloric content), craft beers (mostly only local and produced in limited amounts), non-alcoholic beers (suitable for drivers) and premium beers (representing the mainstream beer). Segments are set according to the market segmentation common in the beer brewing industry. Individual beer brands were assign to each segment according to marketing information of the producers.[9]

The bottom level of our model takes the following form:

$$w_{int} = \alpha_{in} + \beta_i \ln(Y_{mnt}/P_{mnt}) + \sum_{j=1}^{I} \gamma_{ij} \ln p_{jnt} + \epsilon_{int}, \tag{1}$$

$$i = 1, \ldots, I, \quad n = 1, \ldots, N, \quad t = 1, \ldots, T, \quad m \in \{1, \ldots, M\}$$

where w_{int} is the expenditure share of the ith brand from total segment expenditures in the store n and week t, Y_{mnt} denotes the overall expenditure of respective segment m, P_{mnt} is the price index and p_{jnt} is the price of the jth brand. To avoid non-linear estimation when using translog price index, we use lagged Stone index[10] defined as:

$$\ln P_{mnt} = \sum_{i=1}^{I} w_{in(t-1)} \ln p_{int}. \tag{2}$$

Upper level of the model, describing expenditure allocation between the segments, is formulated in similar manner:

$$w_{mnt} = \alpha_{mn} + \beta_m \ln(Y_{bnt}/P_{bnt}) + \sum_{k=1}^{M} \gamma_{mk} \ln P_{knt} + \epsilon_{mnt}, \tag{3}$$

$$m = 1, \ldots, M, \quad n = 1, \ldots, N, \quad t = 1, \ldots, T$$

where w_{mnt} denotes the expenditure share of the mth segment from the overall beer expenditures, whose real values is given by the term Y_{bnt}/P_{bnt}. P_{knt} refers to the price index of the kth beer segment in store n and week t .

We impose restrictions to ensure the sum of all expenditures is equal to 1 (4), segment expenditures are homogeneous in prices (5) and Slutsky's substitution matrix is symmetric (6).

$$\sum_{l=1}^{I(M)} \alpha_{ln} = 1 \qquad \sum_{l=1}^{I(M)} \gamma_{lj(k)} = 0 \qquad \sum_{l=1}^{I(M)} \beta_l = 0 \tag{4}$$

[7] Off-trade market contains beer sales where the sale of beer and its consumption take place at different locations. Off-trade mostly comprises of beer sales in supermarkets and retail stores. On-trade market is represented mostly by beer sales in restaurants and hotels. On-trade and off-trade markets constitute different relevant markets due to the differences in distribution, packaging and existence of buyer power on the off-trade market. It can be assumed that consumer preferences do not significantly vary between on-trade and off-trade market.

[8] This means that the change in price of one brand in a certain segment affects the demand for all brands in other segments in the same way.

[9] Useful cross-check information was also acquired from [14].

[10] As suggested by [5], we use lagged expenditure shares to calculate the price index to avoid simultaneity bias leading to inconsistent estimates.

$$\forall n \in \{1,\ldots,N\}, \quad \forall j \in \{1,\ldots,I\}, \quad \forall k \in \{1,\ldots,M\}$$

$$\sum_{j(k)=1}^{I(M)} \gamma_{lj(k)} = 0 \qquad \forall l \in \{1,\ldots,I\}(\{1,\ldots,M\}) \tag{5}$$

$$\gamma_{lj(k)} = \gamma_{j(k)l} \qquad \forall l,j,k. \tag{6}$$

Elasticities for individual beer brands and beer segments were calculated as

$$\varepsilon_{ij} = -\delta_{ij} + \frac{\gamma_{ij}}{w_i} - \beta_i \frac{w_j}{w_i}, \tag{7}$$

where δ_{ij} denotes Kronecker delta. We use a simplified formula which assume that $w_j = \partial \ln P / \partial \ln p_j$. As shown in [8], this simplification does not change the resulting elasticities.

3 Data

For estimation of our model we use scanner data from U.S. (Chicago metropolitan area) dated from June 1991 until November 1995.[11] Our final dataset consists of panel data on sales volume and prices of individual beer brands in 36 retail chain stores during 220 weeks. All data come from Dominick's Finer Food company, the second biggest retail chain in the Chicago area at that time. Even though the data are not very recent, vast majority of analysed beers is still being sold today. Since consumer substitution pattern has not shown any significant variation during the 220 weeks under investigation, results of our analysis are most likely valid even past the analysed time frame. In order to capture the substitution effects between selected beer brands, we use only data on 6-pack beers abstracting from the substitution effect between different sizes of the same beer brand.

We have detrended and seasonally adjusted the data from weather-induced effect on beer consumption by using average outside temperature.[12] A strong influence on beer sales from several national holidays (Memorial Day, Father's Day, Independence Day, Labor Day, Thanksgiving, Christmas) and sporting events (Super bowl) has also been captured.

4 Estimation

For the model estimation we use fixed effect SUR estimator.[13] SUR estimator allows to include cross-equations constraints (5) and (6) and also captures covariances of cross-equations disturbances, which were significant in our model. When estimating a demand system, one should always check for the endogeneity and consider using instrumental estimators if necessary. Some authors (e.g. [9]) use Hausman-Taylor estimator, which only solves the endogeneity of store specific unobservable effect. Since our data come from retail chain in geographically small area, such effect turned out very small. Only demand shocks across all stores were a potential problem in our case.

We have conducted Hausman tests with estimates from IV 2SLS estimator using lagged own price and wholesale costs as instruments. In most cases, the test rejected the null hypothesis and recommended using the IV estimator. However lack of appropriate instrumental variables did not allow us to use IV estimation. On the other hand, since prices were mostly set independently of demand shocks, potential endogeneity was not considered a very significant problem.[14]

At first we estimate the bottom level of our model. Tabulating full estimation results would be too excessive, so we only briefly comment on them.[15]. The own-price elasticities of almost all beer brands

[11] We would like to hereby acknowledge the James M. Kilts Center, Booth School of Business at University of Chicago that has provided us with the data.

[12] Data on outside temperature were acquired from National Oceanic and Atmospheric Administration (NOAA). Weakly mean temperature at Chicago O'Hare airport has been calculated from hour-by-hour daily temperature data.

[13] Our estimator is a special case of SUR-GLS estimator using only within data variability, for more information see [1, chapter 6].

[14] This is in line with [2], who have analysed the pricing decisions of managers on the same data as we use. Meza (see [11]) (using the same data as well) also shows that retail prices fall in the periods of high demand in order to attract more customers, who once in store also buy other goods with higher margin.

[15] Full estimation results are available on request.

falls in the interval $(-4, -2)$, only one non-alcoholic beer shows inelastic demand curve. Overall, brands from premium and imported beer segments have the most elastic demands, which suggests quite small brand loyalty within the segments. This is surprising esp. with the imported beers, as data show slightly higher demand for German beers during German national holidays. One would thus expect rather high brand loyalty. The least elastic demand have been found for non-alcoholic beers, which is in line with our expectation (non-alcoholic beers are mostly consumed by drivers and hence not substitutable by regular alcoholic beers). Low elasticity is also reported for dark beers, which probably shows high valuation of customers for taste, alcohol content and other unique attributes of dark beers. The demand elasticities for every single beer are quite high to render any beer a relevant market in itself. Thus we focus mainly on the elasticities of beer segments and market delineation on the level of beer segments.

The estimated coefficients, standard errors and elasticities for the upper level of the model are presented in Table 1 [16].

	Premium	**Light**	**Dark**	**Imported**	**Non-alc.**	**Craft**
$\ln(Y/P)$	-0.006*	0.004*	0.006*	-0.004*	0.009*	0.010*
s.d.	(0.002)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)
$\ln(P_{\text{Premium}})$	-0.102*	0.050*	0.007*	0.010*	0.002	0.033*
s.d.	(0.010)	(0.005)	(0.003)	(0.004)	(0.005)	(0.004)
$\varepsilon_{\text{Premium}}$	-1.218	0.110	0.017	0.024	0.005	0.074
$\ln(\text{P}_{\text{Light}})$	0.050*	-0.115*	0.016*	0.028*	-0.013*	0.034*
s.d.	(0.005)	(0.005)	(0.002)	(0.002)	(0.003)	(0.003)
$\varepsilon_{\text{Light}}$	0.374	-1.902	0.125	0.213	-0.105	0.266
$\ln(\text{P}_{\text{Dark}})$	0.007*	0.016*	0.017*	-0.007*	-0.002	-0.032*
s.d.	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$\varepsilon_{\text{Dark}}$	0.062	0.222	-0.760	-0.113	-0.039	-0.465
$\ln(P_{\text{Imported}})$	0.010*	0.028*	-0.007*	-0.004	-0.012*	-0.015*
s.d.	(0.004)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
$\varepsilon_{\text{Imported}}$	0.097	0.222	-0.053	-1.028	-0.091	-0.112
$\ln(\text{P}_{\text{Non-alc.}})$	0.002	-0.013*	-0.002	-0.012*	0.028*	-0.003
s.d.	(0.005)	(0.003)	(0.002)	(0.002)	(0.004)	(0.002)
$\varepsilon_{\text{Non-alc.}}$	-0.024	-0.134	-0.025	-0.125	-0.747	-0.034
$\ln(\text{P}_{\text{Craft}})$	0.033*	0.034*	-0.032*	-0.015*	-0.003	-0.019*
s.d.	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
$\varepsilon_{\text{Craft}}$	0.254	0.295	-0.288	-0.142	-0.032	-1.176
Constant	0.167*	0.167*	0.167*	0.167*	0.167*	0.167*
s.d.	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)

Table 1 Segment level estimation

The statistical significance of the parameters is smaller relative to the inter-segment estimations, but the majority of coefficients still remains significant. The level of the own-price elasticities is in line with our bottom level estimates. The premium and light beer segments show the highest price elasticity of demand, whereas non-alcoholic and dark beer segments report low own-price elasticities. Interesting are negative cross-price elasticities with the non-alcoholic segment. One wonders whether regular beers, which are bought for home celebrations, are really complements to non-alcoholic beers that might be bought for visitors arriving to the celebration by car.

5 Critical Loss Analysis

In order to quantify the Hypothetical Monopolist Test, we use Critical Loss Analysis (CLA). CLA calculates the level of sales that a hypothetical monopolist could lose due to a 5-10 % price increase in order

[16] Symbol * indicates statistically significant estimates at least on 5% significance level.

to preserve at least the level of profit before the price increase. For more information on Critical Loss Analysis see e.g. [3]. Price elasticity of demand, under which the original level of profit (i.e. before the price increase) is equal to the level of profit after the SSNIP is called *break-even critical demand elasticity* ($\varepsilon_{c,be}$). Its level is compared to the actual estimated demand elasticity (ε) for analysed set of products. If $\varepsilon > \varepsilon_{c,be}$, then the set of products does not constitute a relevant market as a close enough substitute exists making the actual demand more elastic (than required by the critical loss elasticity) and rendering the SSNIP not profitable. If the opposite is true, the tested set of products is the relevant market.

A second version of CLA exists - the profit-maximizing critical loss analysis.[17] Under this version of CLA, the question is whether a profit-maximizing monopolist would increase a price by at least 5-10 % in order to achieve a profit-maximizing price. Critical demand elasticity is equal to the elasticity of demand, under which the hypothetical monopolist would increase its price by at least 5-10 % in order to set its price on profit-maximizing level [18].

Since both approaches are performed by the competition authorities when using CLA, we have also performed both types of critical loss analysis. In this section, we assume constant marginal costs equal to wholesale price of each beer brand. We also assume linear demand curve, which is in line with AIDS being first-order approximation of any demand system. The following formulas ((8), (9)) describe the calculation of the break-even critical elasticity and profit-maximizing critical elasticity, where m is initial profit-margin and t is the level of price increase.[19]

$$\varepsilon_{c,be} = \frac{1}{m+t} \qquad (8) \qquad\qquad \varepsilon_{c,pm} = \frac{1}{m+2t} \qquad (9)$$

Table 2 shows profit-margins, estimated elasticities for the segments and the level of break-even and profit-maximizing critical demand elasticity for price increase of 5 % and 10%.

			SSNIP = 0.05		**SSNIP = 0.1**	
	m	ε	$\varepsilon_{c,be}$	$\varepsilon_{c,pm}$	$\varepsilon_{c,be}$	$\varepsilon_{c,pm}$
Premium	0.175	1.218	4.45	3.64	3.64	2.67
Light	0.161	1.902	4.74	3.83	3.83	2.77
Dark	0.108	0.760	6.31	4.80	4.80	3.24
Imported	0.084	1.028	7.47	5.44	5.44	3.52
Non-alcoholic	0.185	0.747	4.26	3.51	3.51	2.60
Craft	0.150	1.176	5.00	4.00	4.00	2.86

Table 2 Profit-margin, estimated elasticity and critical demand elasticities

As can be seen, $\varepsilon < \varepsilon_{c,pm} < \varepsilon_{c,be}$ for all the estimated segments and a given increase in price. It can be concluded, that irrelevant of whether the profit-maximizing or the break-even version of CLA is used, proposed beer segments constitute separate relevant markets. A hypothetical monopolist over each segment could profitably increase its price by 5-10 % (or such price increase would be profit maximizing).

6 Conclusion

In the last ten years, the increasing concentration of the beer brewing industry also meant an increasing number of merger cases assessed by national competition authorities and the European Commission. As the first and very crucial step in every competition case is the definition of a relevant market, competition authorities have been faced with the question whether beer is a relevant market in itself or whether narrower relevant product markets exist. Increasing concentration of the beer brewing market puts more pressure on correct relevant market definition in order to estimate the price effects of the merger.

We have used LA-AIDS model to estimate the own-price elasticities for selected beer segments and conducted Critical Loss Analysis to find out, whether a relevant market narrower than beer exists. Our

[17] For more informations on the difference between the break-even critical loss analysis and profit-maximizing critical loss analysis, see [12].

[18] For more information on Critical Loss Analysis, see [10]

[19] Detailed derivation of the formulas and more detailed intuition can be found in [15].

results show that a hypothetical monopolist over any one of our proposed segments would have enough market power to implement the SSNIP, making each of the analysed segments a separate relevant market. Relevant market in the area of beer brewing thus seems narrower than all of beer and a merger between two beer producers with overall small market share but which both focus on a certain beer segment, might constitute a competition concern.

Our results on the size of the relevant product market are deduced solely on the estimation of the demand function. Further research should be done on potential supply side substitution, likelihood of entry and existence of buyer power - all of which can render SSNIP unprofitable and affect the definition of the relevant market. Analysis of these areas was beyond the scope of this article and will be looked into in the future. Regardless of the shortcomings, demand estimation still remains a crucial first step in every relevant market definition.

Acknowledgements

This work was supported by Student Project Grant MUNI/A/0796/2011.

References

[1] Baltagi, B. H.: *Econometric Analysis of Panel Data*. John Wiley & Sons, 2008.

[2] Bronnenberg, B., Mela, C., and Boulding, W.: The Periodicity of Pricing, *Journal of Marketing Research 43* **3** (2006), 477–493.

[3] Davis, P., and Garces, E.: *Quantitative techniques for competition and antitrust analysis*. Princeton University Press, 2010.

[4] Deaton, A., and Muellbauer, J.: An Almost Ideal Demand System, *The American Economic Review 70* **3** (1980), 312–326.

[5] Eales, S. J., and Unnevehr, L. J.: Demand for Beef and Chicken Products: Separability and Structural Change, *American Journal of Agricultural Economics 70* **3** (1988), 521–532.

[6] European Commission: Notice on the Definition of Relevant Market For The Purposes of Community Competition Law, *Official Journal of the European Communities 1209* **97** (1997), 1–12.

[7] Gokhale, J., and Tremblay, V.: Competition and Price Wars in the U.S. Brewing Industry, *forthcoming* (2011).

[8] Green, R., and Alston, J.: Elasticities in AIDS Models, *American Journal of Agricultural Economics 72* **2** (1990), 442–445.

[9] Hausman, J., Leonard, G., and Zona, D.: Competitive Analysis with Differenciated Products, *Annals of Economics and Statistics / Annales d'Économie et de Statistique* **34** (1994), 159–180.

[10] Johnson, F. I.: Market Definition Under the Merger Guidelines: Critical Demand Elasticities , *mimeo* (1986), 1–13.

[11] Mezza, S., and Sudhir, K.: Pass-through timing, *Quantitative Marketing and Economics 4* **4** (2006), 351–382.

[12] Niels, G.: "Could" or "would"? The difference between two hypothetical monopolists , *Oxera Agenda* **1** (2008), 1–15.

[13] Rojas, Ch., and Peterson, E. B.: *Demand for differentiated products: Price and advertising evidence from the U.S. beer market 26* **1** (2008), 288–307.

[14] Subhash, J.: Global Competitivness in the Beer Industry: A Case Study , *University of Connecticut, Food Marketing Policy Center Working paper* (1994), 1–36.

[15] Werden, G. J.: Demand Elasticities in Antitrust Analysis, *Antitrust Law Journal* **66** (1997), 363–414.

Mathematical Modeling of Economic Phenomena with Maple

Jiří Hřebíček[1]

Abstract. The paper presents new possibilities of Maple 16 for mathematical modeling and interactive solving economic phenomena. There is shortly introduced Clickable Mathematics and concept of its usage by students, teachers and researches in economic modeling. Educational demonstration worksheets of Maple 16 for solving simple economic phenomena are introduced.

Keywords: economic phenomena, mathematical modeling, Maple, Clickable Mathematics

JEL Classification: C63
AMS Classification: 68N01

1 Introduction

Mathematical software Maple [2] is the result of the Canadian company Maplesoft[2] over thirty years of cutting-edge research in computer algebra. This is also user friendly environment for doing mathematics and mathematical modeling, [4], [6], [15] including economics education [7], [8], [17].

Economics is a social science. Its subject matter is society and behavior of institutions and individuals of which it is composed. It is called a science to reflect the fact that, as a discipline, knowledge is built via the scientific method. This simply means that theories about phenomena in the economic domain are subjected to scrutiny that is both logical and empirical. Thus, economists are often called on by those in decision making positions, in government and the private sector, to provide economic input into the decision-making process [7]. The basis for economical analyses at the decision-making level is focused on mathematical modeling of real economical phenomena. Modeling can have miscellaneous features following from differing approach, application of numerous methods or availability of means [18], [23].

Maple helps to analyze, explore, visualize, and solve mathematical problems in economics [18], [23] and finance engineering [21]. The Finance package in Maple contains many tools for advanced financial modeling, as well as accessible tools for personal finance [2]. On the personal finance side, there are tools that can be used for computing with mortgages or retirement packages. The Financial Modeling package supports a wide range of stochastic processes used in Financial Engineering. This includes processes for modeling equity prices, mean-reverting processes, pure jump processes, jump diffusions as well as multivariate Ito processes [5]. In addition, the package provides tools for building more complicated processes from simple ones. Users can also create, manipulate, and analyze many types of financial instruments, such as American, Bermudan, and European options and swaptions [1] and several types of bonds; short rate models; term structures of interest rates; and cash flows. The instruments can then be priced using analytic methods, lattice methods or Monte Carlo simulation - all using one of many date arithmetic conventions. Finally, the processes occurring in the Finance package can be visualized in several ways.

All computations in Maple are performed by a computation engine. It is the command processor, which consists of two parts the *kernel* and the *math library*:

- The *kernel* is the core of the Maple computation engine. It contains the essential facilities required to run and interpret Maple programs, and manage data structures. In this guide, the kernel commands are referred to as built-in commands. The Maple kernel also consists of kernel extensions, which are collections of external compiled libraries that are included in Maple to provide low-level programming functionality. These libraries include Basic Linear Algebra Subprograms (BLAS) [3], GNU Multiple Precision (GMP)[3], the NAG® C Library[4], and the C Linear Algebra PACKage (CLAPACK)[5].
- The *math library* contains most of the Maple commands. It includes functionality for numerous mathematical domains, including calculus, linear algebra, number theory, and combinatorics. It also contains commands for numerous other tasks, including importing data into Maple, XML processing, graphics, and translating Maple code to other programming languages. All library commands are implemented in the high-level Maple

[1] Mendel University, Faculty of Business and Economics, Zemědělská 1, 61300 Brno, hrebicek@mendelu.cz.
[2] http://www.maplesoft.com
[3] http://gmplib.org/
[4] http://www.nag.co.uk/numeric/CL/cldescription.asp
[5] http://www.netlib.org/clapack/

programming language, so they can be viewed and modified by users. By learning the Maple programming language, users can create custom programs and packages, and extend the Maple library.

The majority of the mathematical algorithms there are written in the *Maple Programming Language* [2]. Maple users can write programs using the same basic tools which the Maple developers themselves use. Moreover Maple users can easily view most of the code in the Maple library and they can even extend the Maple system, trying their programs in with existing functionality.

New release Maple 16 offers over 5000 functions with the breadth, depth, and performance to handle every type of mathematics [2]. Now, Maple 16 will raise the bar even higher, introducing new, innovative ways to explore mathematics including mathematical modeling of economic phenomena.

Maple's user community is now over two million people in the world. Together with this community Maplesoft has built large collections of Maple worksheets and Maple programs, many of them are freely available on the Maplesoft web[2] at Application Centre[6] for Maple users to reuse or learn from them. There are 153 matching applications in economics and 26 matching applications in finance engineering.

The *Czech Maple User Group*[7] has supported the cooperation of Maplesoft with the Maple users at the Czech and the Slovak Republic. It was established in 1993 and it organized many workshops, seminars and scientific forums for information exchange of experiences of using Maple in education and research [10] - [14], [16], [17], [19].

The paper introduces chosen features of the last versions of Maple (Release16) for teaching economic modeling online and its applications in scientific disciplines, especially in the area of economical/financial modeling. Students learn mathematical modeling as an important interactive support for understanding and presenting solved real economic and financial problems. In this paper is presented chosen basic ideas on process of mathematical modeling of economic phenomena which are demonstrated in two examples using Maple help system and solved Mathematical applications (Math Apps) in practice including Maple visualization tools [2].

2 Interactive solving economic problems in Maple

The basic Maple graphic user interface consists from worksheets [2]. They are files that document how to solve mathematical problems from the fields of mathematics, science, and engineering. Note that Maple provides users with two worksheet interfaces. Worksheets are both interactive and reusable. They can be used to replace calculators, spreadsheet application programs, and programs in languages such as FORTRAN, C, Java, Visual Basic, etc. Specifically, in worksheets Maple users can:

- perform computations;
- manipulate mathematical expressions;
- describe the problem-solving process.

When a user uses Maple to perform computations or manipulate expressions, Maple displays the corresponding results, which the user can use for subsequent processing. The request sending by user to Maple is called *Maple input*, and the result is called *Maple output*. Maple input and output comprise together an execution group, which is the fundamental element of the worksheet.

Maple users enter, compute, and manipulate mathematical expressions of the worksheets by using the powerful Maple mathematical engine. During the problem-solving process, Maple can take the result of one calculation and use it as input for a second calculation. In this way, Maple can solve very complicated mathematical problems in economics.

2.1 Clickable Mathematics

Maple has led the way in making mathematical software easy to use for years. Its intuitive graphic user interface supports multiple styles of interaction, from *Clickable Math*™ tools[8] to a sophisticated programming language [2]. Using the smart document environment provided by Maple, its users can automatically capture all of his technical knowledge in an electronic form that combines calculations, explanatory text and math, graphics, images, sound, and diagrams. Clickable Math 3.0 refers to Maple's remarkable set of user-interface features that makes common mathematical operations as easy as pointing and clicking. For example users can use *Smart Popups*, which are interactive pop up menus that are automatically displayed when users select an expression or individual terms in an expression, see Fig. 1. They provide a choice of options for manipulating user selection,

[6] http://www.maplesoft.com/applications/
[7] http://www.maplesoft.cz
[8] http://www.maplesoft.com/products/maple/demo/player/ClickableMath.aspx

such as plotting, expanding, factoring, substituting trigonometric expressions, etc. Smart Popups work only on the output of a Maple expression or equation.

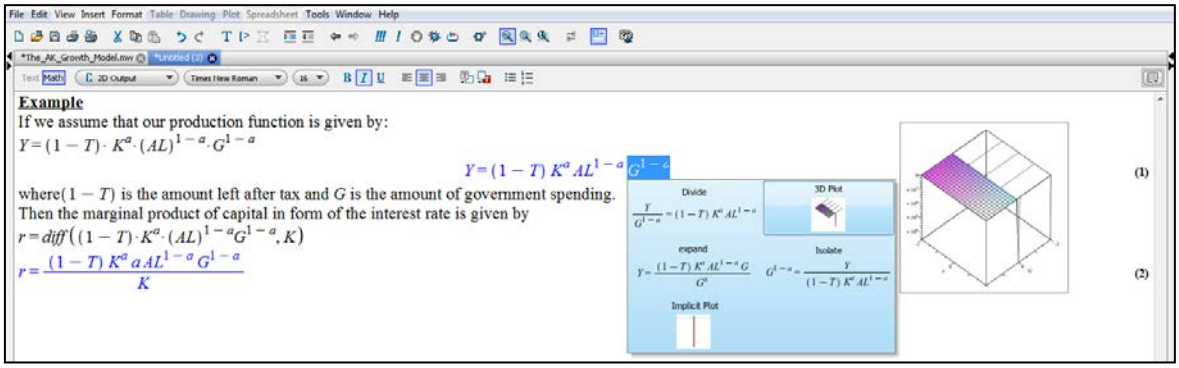


Figure 1 Smart Popups menu in Maple 16

The *Drag-to-Solve* feature of Maple 16 enables users to drag individual terms in an output equation of Maple worksheet from one side of the equal sign to the other. The very simple example is shown in Fig. 2. This action invokes the *Smart Popup* feature which displays user's available options. Using the Drag-to-Solve feature allows users to direct what will happen interactively. It helps with symbolic manipulation.

Input the following expression, then press Enter to produce the required output: $5\cdot x-7=3\cdot x+2$	$5x-7=3x+2$
Drag $3x$ from the right side of the equal sign to the left side. A Smart Popup window is displayed, previewing the results of this manipulation. Click on the Smart Popup window to confirm this operation.	$5x-7=3x+2$ Subtract $2x-7=2$
Next, in the resulting output equation, drag -7 from the left side of the equation to the right side. Click on the Smart Popup window to confirm this operation.	$2x-7=2$ Subtract $2x=9$
Drag and drop the factor of 2, in front of x, to the right side of the equal sign. Click on the Smart Popup window to confirm this operation.	$2x=9$ Divide $x=\frac{9}{2}$

Figure 2 Example of interactive solving the linear equation using Drag-to-Solve

The concept of Clickable Maths for solving economic phenomena includes [17]:

- smart, context-sensitive right-click menus for instant access to Maple solvers of economic phenomena and other command-free operations in mathematical modeling;
- extensive range of Maple palettes for visual editing of mathematical expressions;
- Interactive plotting and animation controlled by the mouse and not by endless parameters and attributes in a Maple command;
- drag and drop and Drag to Solve operations on expressions, plots, text, and more;
- Interactive assistants that provide easy mechanisms to solve and explore advanced topics for economical modeling such as linear, nonlinear and differential equation-solving, optimization, and advanced visualization;
- Maple Portal for Students, which acts as a guide for hundreds of common tasks from courses of mathematics, economics, econometrics, finance engineering, etc;
- built-in selection of interactive tutors that offer visual e-learning environments as for many important mathematical topics in precalculus, calculus, linear algebra, as for economics and economic modeling;
- handwriting recognition of mathematical symbols and equations in economics;
- WYSIWYG document (worksheets) processing features that let Maple users create complex documents of economics courses more quickly and easily than in a word processor or LaTeX;
- Exploration Assistant that allows students and teachers to instantly create interactive mini economic applications and Math Apps examples to explore the parameters of expressions and formulas.

We can summarize Maple provides both computational power and a new-generation user interface that allows students, teachers and researcher to become proficient with Maple without the burden of learning commands and their related syntax. The result is that teachers can spend their time teaching mathematics and solving economic phenomena rather than the Maple commands. Overall, teaching and learning economics become more efficient and effective. With collection of Clickable Math tools, including palettes, interactive assistants, context-sensitive menus, tutors, and more, Maple has set the standard for making it easy to learn, teach, and do mathematics and solve mathematical problems in economics [1], [5], [8], [18], [21], [23] - [26].

3 Educational demonstration worksheets for solving economic problems in Maple

For Maple 16, Maplesoft has added over 100 useful and easy to use educational demonstration worksheets called *Math Apps*, designed to illustrate various mathematical, physical and economic concepts. Each demonstration has a brief section describing the concepts involved and some interactive features that bring the concepts to life and allow students to test their understanding. The student can type mathematical equations, click buttons and adjust sliders, draw, click or drag on the plot itself, and with Maple's powerful engine updating the calculations in the background, the student can see their actions make effect instantaneously. The worksheets can be accessed by choosing Math Apps from the Tools menu of Maple. We present these possibilities in two examples.

Example 1. When everybody first starts paying down his *mortgage*, nearly all of his payment is applied to the interest, with very little going to pay back the principal (also called the equity). With each payment, slightly less goes to interest and slightly more to principal, until the last payment is almost all principals. Four parameters determine the total cost of a mortgage[9]: The *principal (equity)*, the *interest rate*, the *frequency of payments*, and the *amortization period (total time for repayment)*.

We implemented known mathematical formulas of finance mathematics [7], [21] and used the Maple sliders [1], see below in the living graphs of Fig. 3. They enable to change the parameters interactively for the principal ($20000.00), interest rate (5%), payment frequency (12 per year), and period of the mortgage (20 years) and obtain also total costs ($316778.74) which it is necessary to pay during period of the mortgage. The slider components are defined in a Standard Maple worksheet or document [2]. These components perform an action or a series of actions to communicate with other embedded components when altering the slider position in worksheet, although assigning an action to a component is not required [2].

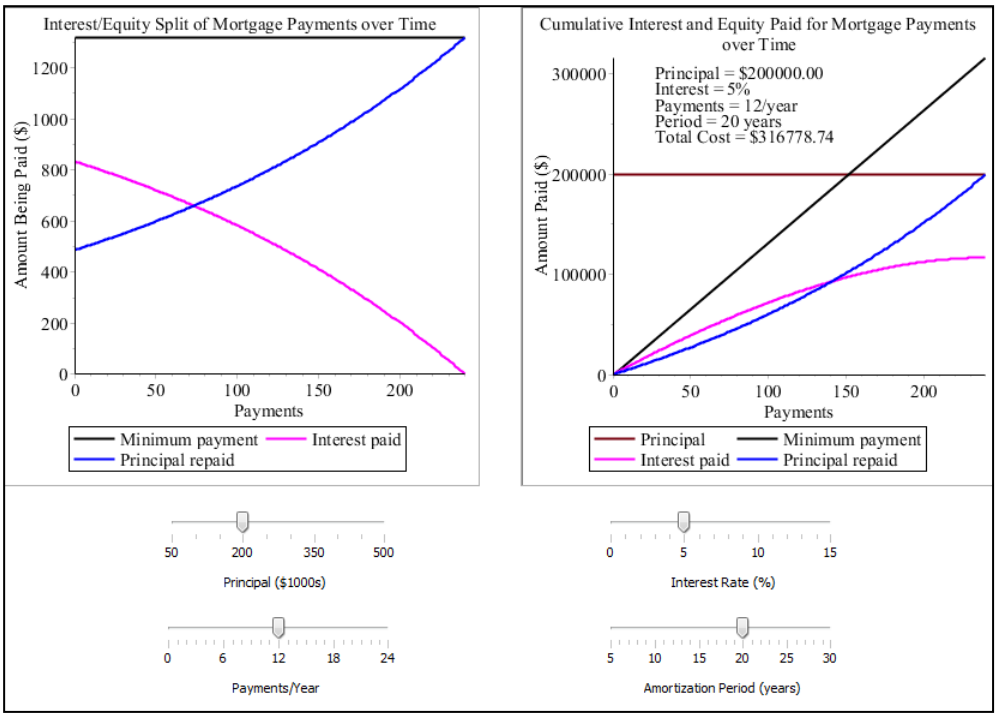


Figure 3 Mortage educational demonstration worksheet

[9] http://www.hughchou.org/calc/formula.html

In the Fig. 3, Maple 16 explores how these parameters affect the structure of the mortgage. The black lines are the minimum payment per payment period (for example the monthly payment), the magenta lines are the amount of interest being paid, the blue lines are the amount being paid off of the loan, and the red line (right graph only) is the principal. The left graph shows the value of these over time, while the right graph shows the accumulated principal and interest paid over time. The magenta and blue curves are exponential functions. The living graphs will be presented during lecture.

Example 2. *Supply* and *demand* are two fundamental concepts of economics. The law of demand states that as price increases, the quantity of items demanded by consumers' decreases [7]. The law of supply states that the higher the price, the greater quantity supplied by the manufacturers. For given supply and demand curves, the market price and quantity of goods sold will fluctuate until they reach *equilibrium* at the point where the two curves intersect. At equilibrium, the number of goods demanded is the same as the number of goods that can be manufactured at that price.

The graph of educational demonstration worksheet at Fig. 4 shows simplified models of the supply (blue) and demand (green) curves. Using the Maple sliders [2] we can adjust the parameters of the curves and observe how their point of intersection changes.

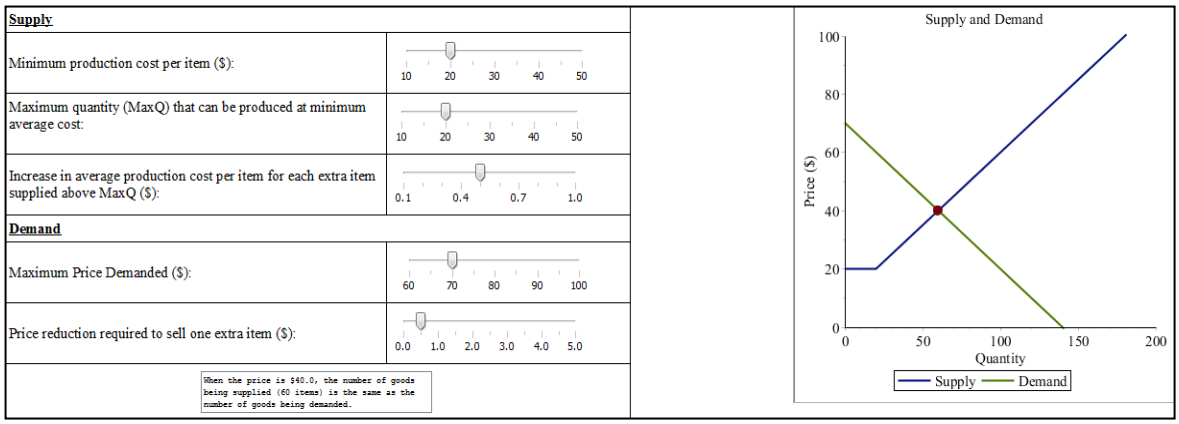


Figure 4 Supply and demand educational demonstration worksheet

4 Conclusion

Mathematical modeling of economic phenomena with using Maple in bachelor and master study courses of economics (mathematics, microeconomics, macroeconomics etc) has led to substantial and meaningful improvements of pedagogical processes at Faculty of Business and Management of Brno University of Technology (FBM BUT) and Faculty of Business and Economics of Mendel University in Brno (FBE MENDELU) [17]:

- Maple has made the tools of mathematical modeling of economic phenomena much easier to students, teachers and researchers.
- The graphical capabilities of Maple have allowed students to explore economic phenomena with much more detail than previously possible.
- Using Maple enables to remove the algebraic tedium of symbolic mathematical manipulation in economic formulas. This has allowed students not "to lose the forest for the trees" but to understand economic background of solved problem. Thank to students have a better appreciation for the mathematical foundation of microeconomics and finance engineering and are able to spend more time thinking about problem. Their economic intuition enables to built their solutions rather than concentrating themselves on the disproportionately calculation of the problem solutions.
- Using the Maple tools to solve mathematical models of economic problems should allow economists to begin to explore a larger variety of utility functions and other mathematical models of economic phenomena.

Interactive Maple tools are also used for solving more difficult economic problems in the common research project No.P403/11/2085 of the FBM BUT and FBE MENDELU [20], but there is not more place to present them here.

Acknowledgements

This paper is supported by the Czech Science Foundation. Name of the Project: Construction of Methods for Multifactor Assessment of Company Complex Performance in Selected Sectors. Reg. No. P403/11/2085.

References

[1] Andersen, L. and Broadie, M.: Primal-Dual Simulation Algorithm for Pricing Multidimensional American Options, *Management Science* **50** (2004), 1222-1234.

[2] Bernardin, L., Chin, P., DeMarco, P., Geddes, K. O., Hare, D. E. G., Heal, K.M., Labahn, G., May, J. P., McCarron, J., Monagan, M. B., Ohashi, D., and Vorkoetter, S. M.: *Maple 16 Programming Guide*, Maplesoft, Waterloo, 2012.

[3] Dongarra, J.: Basic Linear Algebra Subprograms Technical (Blast) Forum Standard (1), (2). *International Journal of High Performance Computing Applications* **16** (2002), 1-111, 115-199.

[4] Gander W., and Hřebíček J.: *Solving Problems in Scientific Computing Using Maple and Matlab*. 4th expanded and rev. ed., Springer, Heidelberg, 2004.

[5] Glasserman, P.: *Monte Carlo Methods in Financial Engineering*, Springer, New York, 2004.

[6] Gregor, J., and Tišer, J.: *Discovering Mathematics: A Problem-Solving Approach to Mathematical Analysis with MATHEMATICA® and Maple™*. Springer, Heidelberg, 2010.

[7] Griffiths, W. E.: *Learning and Practicing Econometrics*. John Wiley & Sons Inc., New York, 1993.

[8] Hey, J. D.: *Intermediate Microeconomics: People are Different*, Pearson Custom Publishing, 2011.

[9] Hřebíček, J., and Chvátalová, Z.: Computer Simulation of Stock Exchange Behavior in Maple system. In: *Proceedings of 9th International Conference: Applied Computer Science* (Deo, N., et al., eds.). WSEAS Press, Malta, 2010, 432–436.

[10] Hřebíček, J., and Chvátalová, Z.: Chosen Features of Maple for Mathematical Modeling in Education and Practice. *Journal of Applied Mathematics* **4** (2011), 379-393.

[11] Hřebíček, J., and Chvátalová, Z.: Maple as an Instrument for Connection School and Practice. In: *Proceedings of 8th International Conference on Applied Mathematics (APLIMAT 2009)* (Kováčová, M., ed.). Slovak University of Technology, Bratislava, 2009, 667-681.

[12] Hřebíček, J., and Chvátalová, Z.: Quantitative methods in economic computation using Maple (in Czech). In: *Proceedings of the 5th International Conference APLIMAT 2006*. (Kováčová, M., ed.). Slovak University of Technology, Bratislava, 2006, 567-606.

[13] Hřebíček, J., and Řezáč, M.: Modelling with Maple and MapleSim. In: *Proceedings of the 22nd European Conference on Modelling nad Simulation ECMS 2008*, Digitaldruck Pirrot, Dudweiler, 2008, 60-66.

[14] Hřebíček, J., Chvátalová, Z., and Vecheta, L.: Modelling the economic indicators using Maple. *Systémová integrace* **4** (2010), 27-42.

[15] Hřebíček, J., Pospíšil, Z., and Urbánek, J.: *Introduction to mathematical modeling with using Maple* (in Czech). Masarykova univerzita, Brno, 2011. Online <http://is.muni.cz/do/rect/el/estud/prif/js11/maple/web/mat_model.pdf>

[16] Hřebíček, J.: *Teaching mathematics with using Maple* (in Czech). In: *Proceedings of the 5th International Conference APLIMAT 2006*. (Kováčová, M., ed.). Slovak University of Technology, Bratislava, 2007, 357-362.

[17] Chvátalová, Z., and Hřebíček, J.: Education of Economics with Maple. In: *Proceedings of the International Conference on Environment, Economics, Energy, Devices, Systems, Communications, Computers, Pure and Applied Mathematics* (Lazard M. Et al., eds.). WSEAS Press, Florence, 2011, 435-440.

[18] Chvátalová, Z., and Šimberová, I.: Economic phenomena via mathematical modeling in Maple system, *Business, Management and Education* **9** (2011), 260–276.

[19] Chvátalová, Z., Hřebíček, J., and Žigárdy, M.: Computer simulation of stock exchange behavior in Maple. *International journal of mathematical models and methods in applied sciences* **5** (2011), 59-66.

[20] Chvátalová, Z., Kocmanová, A., Dočekalová, M.: Corporate Sustainability Reporting and Measuring Corporate Performance, In*: Proceedings Environmental Software Systems. Frameworks of eEnvironment. 9th IFIP WG 5.11 International Symposium. ISESS 2011*. (Hřebíček, J. et al, eds.) Springer, Heidelberg, 2011, 398–406.

[21] Lynch, S.: *Dynamical Systems with Applications using Maple*. Birkhauser/Springer, New York, 2009.

[22] Řezáč, M.: Estimating Information Value for Credit Scoring Models. *Journal of Applied Mathematics* **4** (2011), 1619-1628.

[23] Shone, R.: *Economic Dynamics. Phase Diagrams and Their Economic Application*. Cambridge University Press, 2002.

[24] Siddiqi, A. F.: Financial Engineering Computing: Binomial Modeling Using Maple. *Proceedings Pakistan Academy of Science* **43** (2006), 211-220.

[25] Snarr, H. W., and Gold, S. C.: The design and use of macroeconomics simulation using Maple software: A pilot study. *Developments in Business Simulation and Experiential Learning* **33** (2006), 253-262.

[26] Soukopová, J., and Bakoš, E.: Assessing the efficiency of municipal expenditures regarding environmental protection. In: *Proceedings of Environmental Economics and Investment Assessment III* (Aravossis, K., and Brebbia, C.A., eds.) WIT Press, Cyprus, 2010, 107-119.

ARIMA model selection in Matlab

Radek Hřebík[1], Jana Sekničková[2]

Abstract. This paper aims to discuss and suggest an approach to analyzing and modelling of economic time series. Econometric theory deals with the problem of right models. As time series analysis methodology is selected the Box-Jenkins methodology representing the stochastic approach. Aim of this paper is to propose an interactive application to enable user not only automated selection of time series model but also to inform the user about everything important that has been done in process of automated selection. This paper deals with Matlab as one of not very typical tool for construction of time series models. Matlab was selected to show and emphasize the power of such tool commonly used at technical universities. The tool gives to user a wide range of possibilities to assess the model. We suggest an application which enables automated model selection and works in two modes, for basic and advanced user. Aim of two user groups is not to discriminate the beginners and not to bore the advanced users. The automated model selection is planned to be divided into two phases. The first phase includes model identification and the second is based on model verification.

Keywords: ARIMA, Box-Jenkins, model selection, Matlab.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The time series analysis and modelling represent important process and are very often demanded in various areas of life. The reason is simple because in most cases it is needed to predict future values of time series. The answer to why predict future values is quite clear. Almost everybody wants to know something about the future progress, about the future opportunities. In some fields it may be also the main content to predict future values. As example illustrating this claim can serve the compilation of public budgets. Budget for next period is almost always based on the prediction of main indicators. Econometric theory gives answer to this and the existence of econometric models and their usefulness is nowadays almost doubtless. But as it always is, nothing is fully ideal. There is high probability that the right model exists, but how to choose them? Number of models that are available today is huge. Surely exist the ways how to select the model, but they are not automated and require experience of users. Moreover, the model selection starts already with the selection of an analysis approach.

Current theory offers two main approaches to time series analysis and model. The selection is between two main approaches – the deterministic and stochastic approach. The deterministic approach assumes that all the facts affecting the time series can be explained exactly, for example with some existing models or so on. In case of stochastic approach there is calculated with a random component. As time series analysis methodology used by authors of this paper is selected the Box-Jenkins methodology representing the stochastic approach.

In 1970 Box and Jenkins made autoregressive integrated moving average (ARIMA) models very popular by proposing a model building methodology with three steps – model identification, estimation of parameters and diagnostic checking (Box and Jenkins [1] and Box, Jenkins and MacGrego [2]) and using obtained model for forecasting. Unfortunately, building an ARIMA model is often a difficult task for users because it requires good statistical practice, knowledge in the field of application and very specialized user-friendly software for time series modelling. There exist a lot of freeware or shareware econometric tools helping users to analyze and model time series. The main aim of such tools is to create econometric model specified by user and inform user about the basic parameters of created model. In many cases the user has to make the decision which model to select and what more, the user is responsible for model verification. Because of using this methodology the model selection is focused on suggestion of some approach to select the right model representing the concrete (ARIMA) process.

[1] Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering, Department of Software Engineering in Economics, Břehová 7, Praha 1, 115 19, Czech Republic, Radek.Hrebik@seznam.cz.
[2] University of Economics, Prague, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 4, 130 67, Praha, Czech Republic, Jana.Seknickova@vse.cz.

Another problem relates with the fact that the number of time series to be analysed is often large and so Box-Jenkins methodology requires both experience and time for successful modelling. There are reasons why we discuss the problem of automated time series analysis. But we are not the first. Also Hoglund and Ostermark in [4] presented study about automatic ARIMA modelling. Unfortunately, authors worked with computer program ANSECH (Mélard [9]) that is not so popular as the others. Tashman and Leach [11] published the article about automatic forecast software but in this work was presented automatic forecasting and not entire modelling process with analysis.

Although the model selection problem is nothing new, it has still no exact solution how to select the right model without human interaction. This paper aims not only to discuss and suggest some solution in this field but aims also to serve as an introductory paper to other possible developers of such application. Application for automated selection discussed in this paper is based on using not typical programming environment to time series analysis. The reason for the aim of this paper is to start a discussion and to inspire the reader to their contribution to solving the automated selection problem. As in most cases it is not possible to select only one model, the offer of more suitable models is also discussed.

What we aim is to propose an interactive application to enable user not only automated selection of time series model but also to inform the user about everything important that has been done. The application aim is also to enable user to suggest his model that was not suggested by our application. The model parameters have to be estimated. In terms of parameter estimation would be used only the implemented methods in existing software (Matlab). It is not the aim of this paper to focus on methods of parameter estimations. To verify the model will be also used methods and functions implemented in used software.

There already exists a lot papers talking about the theory of ARIMA processes, such as Emmenegger [3] or Makridakis and Hibon [8], and about the right models for something like typical or ideal situations but reality seems quite different. In theory is described the ideal process but sometimes it is quite difficult to apply it to the real data. Papers presenting some approaches to these examples from real life which do not fit the model as described in theory are rare. Almost always there is some appeal on analytic decision based on his or her experience. Situation when the data do not fit the model exactly is very common and without human interaction is very hard to select right model to be estimated. The reason of writing paper trying to automate the ARIMA model selection is simple because the question of automated selection is not closed yet.

2 ARIMA and Matlab

In this part of the paper we focused on the selecting criteria for the ARIMA model. As already said the problem of automated model selection is nothing new, but our asset we see in such discussion connected with Matlab and in range of planned use of our application. This paper does not aim to give some genial and cure-all approach. The aim is also to inveigle the reader to his own work and self-evident to show and discuss some example of such approach.

This paper deals with one of not very typical tool for construction of time series models. For model selection, estimation and verification is used high-level computing language Matlab. The reason for using Matlab is implemented suitable econometric tool for parameter estimation helping us in our decisions how to select the right fitting ARIMA model. Matlab is often used for time series modelling, e.g. in Kugiumtzis and Tsimpiris [7] or in Peng and Aston [10].

The reason why we have selected Matlab is to show and emphasize the power of such tool commonly used at technical universities where is often available for students. Such tool gives the user a wide range of possibilities to construct and assess the model. The strength of tool as Matlab and aim of this paper shows one of the reasons why not to use a typical econometric software. Using typical econometric software is nothing bad and in practise it is also seen as a best way. But if the user wants to work more with the constructed models and wants to try some own improvement in the common econometric software is not permitted to do some changes. Here we see the biggest advantage of tool, such as Matlab, to our purposes. The implementation of used functions to evaluate basic characteristics of time series is available to user. So not only to program own functions in Matlab but in many cases is also possible to improve the existing functions. Because there are many of provided functions looking the same as user created functions.

Because the authors are affiliated with two universities it is also at this point possible to discuss and compare Matlab with other software used in special courses focused on time series modelling. In case of Czech Technical University there are two special courses on econometric and applied econometric. In both courses are students working with software to analyse and model time series. As starting program is in this courses used Gretl, but because the students already have experience with Matlab from other courses, they very often use the Matlab instead of Gretl. The example of using Matlab is also included in the course syllabus. When we mention Univer-

sity of Economics in Prague, there are more courses focused on econometric field. But in fact in the courses which are similar in content we have to conclude that the number of students using Matlab is not as high as in case of Czech Technical University in Prague but is increasing in recent years. Using Matlab gives more opportunity to study the problem [5].

Selecting Matlab to analyse time series using the Box-Jenkins methodology is a very pleasant way. It was already mentioned the use of Matlab implemented functions to estimate the model parameters. The reason why we are able to use functions to estimating parameters is the econometrics toolbox. This toolbox provides functions for modeling economic data. User is able not only to select model to be estimated, but also simulate and forecast. Unused in our case of use to automated model selection, but important to mention is, that toolbox provides for example also Monte Carlo methods for simulating systems of linear and nonlinear stochastic differential equations. Of course the toolbox offer a variety of diagnostics for model selection, including hypothesis, unit root (in Matlab: `h = adftest(Y,Name,Value)`), and stationarity tests (in Matlab e.g.: `h = kpsstest(y,Name,Value,...)`).

2.1 Application user modes

Our suggested application to enable automated model selection is working in two modes. We decided to this division because we want to offer an application with possibly the widest range of use. One application user mode is set for basic user and second for advanced user. The aim is to extend the number of possible users and especially not to discriminate or unnecessarily discourage the beginners. In the case of advanced users we do not want to bore them. Our aim is to enable advanced user the intervention into process of building the right time series model.

The main difference between the two types of users is that the basic user is not informed about all the results of partial processes. Because the interactive user is welcome, such user in advanced mode is able to select whether he or she wants to step in at some checkpoints or not. So our suggested application offers menu to advanced user and there it gives the possibility to select the partial processes.

In case of stepping in there will be as output the selected model with the detailed characteristics for the selection. Because our aim is to get some relevant feedback about successfulness of our application in real use the user will not be able to refuse suggested models. The evaluation of suggested models is very important to the future assessment of successfulness. Although the user is not able to refuse suggested model but because the interactive application is he or she able to suggest some own ARIMA model which lacks in the list of suggested models. When user is not interested in this selection application will continue and give all the steps as output at the end.

Because the application is enabling user to input own suggestion of ARIMA model it is very suitable to uncover the behaviour of time series which led to the selection. The identification part includes test of stationarity of time-series (unit root test should be implemented), and also calculation of values of autocorrelation and partial autocorrelation function. The selection phases will be discussed in the next, at this place is only necessary to mention this.

So after one or more models have been selected to be estimated (in Matlab: `fit = estimate(model,Y)`) follows the model verification. It is not needed to show the basic user all models but only models that satisfy our criteria. The reason for our approach to not to show some models is, that models not passing the verification are for basic user useless to be discussed in next. Of course this approach goes with the problem of that no model has reached criteria. This case is in our application taken with special care and there is created some report informing about the state of art. The user of application will be informed about the result that no model was found, but will be able to see what models where suggested and why the verification failed. In case of advanced needs the problem of no verified model no special attention because the user is always informed completely. In this phase it is necessary to check constant standard deviation of random element (ARCH test in Matlab: `[h,pValue,stat,cValue] = archtest(res,Name,Value)`), autocorrelation of random element (Ljung–Box Q test in Matlab: `[h,pValue,stat,cValue] = lbqtest(res,Name,Value)`), normal distribution of random element (Kolmogorov-Smirnov test in Matlab: `h = kstest(x,CDF,alpha,type)`), significance of parameters (Matlab results include *t*-values, *t*-test in Matlab: `h = ttest(...,alpha)`) and more models passing verification tests (Akaike information criterion – AIC, Bayesian information criterion – BIC in Matlab: `[AIC,BIC] = aicbic(LLF,NumParams,NumObs)`).

The aim of both user categories is to collect data that give us the possibility to evaluate the successfulness of application. To completely accomplish the aim of developed application for automated selection is needed to select how to quantify the results of estimated models. The reason of such approach to our software is that we want to have some feedback from advanced users. We are conscious that every case of real data time series

analysis is unique process and so the collecting of data we see as the best way how to improve our application. The data separated into two categories of basic and advanced users enable us also to view the difference between the using the application.

One of possible the possible conclusions can be also that the two groups of user are unnecessary. But we think that such conclusion is not possible. We see the potential benefit of two user groups in the evaluation of successful estimated models. Because we are aware that basic users can be using time series not suitable for Box-Jenkins methodology.

2.2 Main selection phases

The automated model selection in our application we suggest to divide into two phases. As already said according to the theory is the analysis process in case of Box-Jenkins methodology composed from three phases – model specification, parameters estimation and model verification. The reason why we mention only two phases instead of the three known from theory is that we focus on our contribution to phases. This is the reason why the estimation is not presented as separate phase.

The first phase includes model identification. The identification is based on automated approach which is presented in this paper. The second phase is based on model verification and will be also performed by our application in automated way.

To identify the degree of differencing is needed to study the autocorrelation function (ACF). Here is the possible automated approach very significant. It is talked about the unit roots and existence of them can be simply identified from ACF values. To decide about the unit roots existence is needed to have the rule saying when the ACF values are closed to one, the decision is based on statistic approach and the differencing is done when there is more than 90% probability that the first values are closed to one.

In case of differencing there is made back control if the differenced time-series really fulfil the stationary. If the results are in this phase not significant there is implemented also the possibility of second differencing. In case of failure of differencing the application tries to make logarithmic transformation as one of the other options to make time series stationary.

After suggestion how to make time series stationary there is another very import decision to select the autoregressive (AR) and moving average (MA) part of ARMA model. In this case are very important autocorrelation function (ACF) and partial autocorrelation function (PACF) values. The theory gives clear description of ACF and PACF in case of typical AR, MA or mixed ARMA processes. To decide about the autoregressive or moving average part of process is needed to research the values on its statistical significance and compare them with typical models.

Matlab provides implemented functions to evaluate ACF (in Matlab: `autocorr(Series,nLags,M,nSTDs)`) and PACF values (in Matlab: `parcorr(Series,nLags,M,nSTDs)`) there is no need to define special functions to get theses values. The decision how to interpret the values of ACF and PACF is the theme of our research and makes the engine of our proposed application. Our suggestion is based on fitting real time series. The end of this phase is the selection of model prepared to be evaluated and tested. Why do this and not so only to evaluate all common used models, evaluate them and then compare? The reason is simple, because the evaluation of model is much more demanding than the evaluation of the right approach to model selection. With the term right approach we mean the approach which would not be more demanding then evaluation of different models.

There is nothing wrong on selecting more models to be evaluated and at this place is our program coming into second phase of use. In this phase will the models be constructed and consecutively verified. As it is seems this phase connects the estimation, fully done by Matlab, and the verification phase. The assessment of models is very important and the question is what the right criteria?

Verification has three basic steps. Autocorrelation of random element, normal distribution and statistical significant of estimated parameters. The aim is to decide what model has passed the validation criteria. Ideal case of every three criteria fulfilled is rather rare. As default we work with 10% probability that the test has not confirmed our expectation. The user in basic mode is able to change the default value but only to all the verification tests as whole. Advanced user is selecting the probability of failing to each test separately. As the tests are used commonly used and recommended tests. As in case of parameters estimation also the verification tests are made using implemented in Matlab.

The suggested application deals with ARCH test to ensure about constant standard deviation, to exclude the autocorrelation is used Ljung–Box Q test and research of normal distribution procures the Kolmogorov-Smirnov test. The question of significant parameters is solved by Matlab when estimating model parameters because the estimation result includes *t*-values which are interpreted by our application at this point.

If after verification tests there are more models passing verification then is nothing better than apply the criteria specially designed to this situation. The mainly used are Akaike information criterion (AIC) and Bayesian information criterion (BIC). Both these criteria, providing model quality measurement, are implemented in Matlab so there is no need to create own functions for the application proposes.

3 Conclusion

In this paper was discussed the automated time series analysis. Paper is focused on Box-Jenkins methodology representing the stochastic approach, and proposes an approach helping user to select the right time-series model to estimate the future values. The aim of the paper was to discuss and suggest an application enabling user the automated selection of ARIMA model. The application uses Matlab so the phase of estimation parameters is fully done using Matlab implemented functions. The authors aim is automated approach to specification and verification of the model.

Application is designed for two groups of users. The first group are the basic users. To them provides the application very easy way to get acquainted with the time series analysis during the fully automated model selection. In case of advanced users is application more interactive. Advanced users are able not only to suggest own ARIMA models to be estimated and verified but also to set parameters for verification. To the possibilities of advanced users also belongs the detailed overview of specification process. As the authors are aware that proposed application does not mean the end of their work they are ready to improve the application in future.

To be able to the future improvement have the authors included the evaluation of successfulness. The application is collecting data on the results of the use. There are collected only data connected with the results of automated selections. The term collect sounds at first sight not very good, but the aim is not to select any data connected with user or their analysed data, the data to collect means only the data about successfulness of our application. To the goal of selecting data is to quantify the successfulness of our application in real use and consecutively to improve our application based on the results of considerable amount of real data. The assessment will be made separately for basic and advanced users. We find the assessment of collected data as the best way to future improvement of our application. The problem also may be users are using wrong data to be fitted on ARIMA model.

The phase of specification includes the decision about stationary of time-series. There is primary used the 90% probability and attention is paid also to standard deviation. In case of autoregressive and moving average part of model is the decision based on ACF and PACF values. The values for the investigated time series are compared with the values for typical models known from theory.

Model verification is default set to 90% probability of accomplishment criteria. Users in basic mode are able to change the probability only as a one number applied to all the verification tests. Advanced users may determine own probabilities for each verification test. Because it is possible that more models will satisfy the verification test, we have in also mentioned the criteria helping to select one of more suitable models which are also implemented in our application.

Because of the experience with special courses focused on econometric at universities see the authors also the following possibility of use. The proposed application can serve for students at these special courses. Because the application is ready to evaluate the results, there is a great chance to get data from closely aimed and specialized groups of users. So in ideal case can be data collected also at other selected universities. The specialized group of user ensures the relevance of obtained data and the chance to possible improvement of our application is rapidly increasing. The reporting of errors can serve not only to quantifying the application successfulness but also, when used as tool for students, to assess what mistakes students are making.

Acknowledgements

This work was supported by grant no. SGS11/166/OHK4/3T/14 at Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering and also by grant no. IGA F4/1/2012 of Inter grant agency at University of Economics, Prague, Faculty of Informatics and Statistics.

References

[1] Box, G. E. P. and Jenkins, G. M. (1968) 'Some Recent Advances in Forecasting and Control I.', *The Royal Statistical Society Series C-Applied Statistics,* Vol. 17, Issue 2, pp. 91.

[2] Box, G. E. P., Jenkins, G. M. and MacGrego, J. F. (1974) 'Some Recent Advances in Forecasting and Control 2.', *Journal of the Royal Statistical Society Series C-Applied Statistics,* Vol. 23, Issue 2, pp. 158-179.

[3] Emmenegger, J.F. (1996) 'Time and frequency analysis of economic time series', *Zeitschrift fur angewandte Mathematik und Mechanik*, Vol. 76, Issue 3, pp. 417-418.

[4] Hoglund, R. and Ostermark, R. (1991) 'Automatic ARIMA Modeling by the Cartesian Search Algorithm', *Journal of Forecasting*, Vol. 10, Issue 5, pp. 465-476.
[5] Hřebík, R., Sekničková, J.: 'Preference for econometric software in tuition', *Proceedings of the 9th International Conference on Efficiency and Responsibility in Education (ERIE 2012)*, Prague, pp. 230-238.
[6] Hřebík, R., Sekničková, J.: 'Automated selection of appropriate time-series model', *Proceedings of Quantitative Methods in Economy 2012 Conference* , May 30 – June 1, Bratislava (Slovak Republic), 2012
[7] Kugiumtzis, D. and Tsimpiris, A. (2010) 'Measures of Analysis of Time Series (MATS): A MATLAB Toolkit for Computation of Multiple Measures on Time Series Data Bases', *Journal of Statistical Software,* Vol. 33, Issue 5, pp. 1-30.
[8] Makridakis, S. and Hibon, M. (1997) 'ARMA models and the Box Jenkins methodology', *Journal of Forecasting*, Vol. 16, pp. 147–163.
[9] Mélard, G. (1982), 'Software for time series analysis', *Proceedings in Computational Statistics*, Vienna, pp. 336–341.
[10] Peng, J. and Aston, J. A. D. (2011) 'The State Space Models Toolbox for MATLAB', *Journal of Statistical Software*, Vol. 41, Issue 6, pp. 1-26.
[11] Tashman, L. J. and Leach, M. L. (1991) 'Automatic forecast software: a survey and evaluation', *International Journal of Forecasting*, Vol. 7, pp. 209–230.

PREFACE

Dear conference participant,

It is a great pleasure to welcome you to the 30th MME 2012 Conference organized by the Czech Society of Operations Research, Czech Econometric Society in cooperation with the Silesian University in Opava, School of Business Administration in Karviná. The conference offers you the opportunity to meet the operations research and econometric community in the Czech Republic and also many researchers coming from Slovakia, Poland and other seven countries. It also offers you an exposure to the evolution of the several areas that compose Operational Research, Econometrics, and generally mathematical methods applied in economics to keep you updated on our continuously evolving dynamic disciplines. During the conference you will certainly meet old and new colleagues, exchange ideas, develop new projects. You will also feel and enjoy the special atmosphere of the eastern Ostrava-Karviná region of the Czech Republic.

Beside the scientific programme which includes more than 190 papers, the social programme is also very rich. Let me invite you particularly to the excursion offered by the organizers to Vítkovice Steel (part of the city of Ostrava), where production of iron, coal and agglomerates in so-called Bottom area (Dolní oblast) has been already terminated. Part of this strategic locality was proclaimed National Cultural Monument together with Hlubina coal mine. Here you can visit the original blast furnace, coal mine tower or the gas storage transformed into modern concert hall and other interesting sites.

This conference Proceedings is divided into two parts and includes 179 papers selected from more than 200 papers submitted to the conference programme committee. All published papers have been subjected to a strict reviewing procedure of two independent referees. A positive feature is that our Proceedings includes 33 papers published exclusively by young researchers – mostly doctoral students and more than 30 papers with young scientists as co-authors.

I am confident that you will find the 30th MME 2012 Conference stimulating, rewarding and pleasant and that you will enjoy your stay in Karviná.

In Karviná, September 2012

Prof. Dr. Jaroslav Ramík
Chair of the programme committee

Relationship between human capital and economic growth: The case of Austria

Silvie Chudárková[1], Tomáš Verner[2]

Abstract. Impact of education, more precisely impact of skills and knowledge (human capital) to increase individual productivity and wealth of a nation, was already emphasized by classical economists. However, their opinions and findings were not included in detailed analyses and systematically processed. Comprehensive concepts of theory of human capital were introduced in the early Sixties of 20th century. In the last twenty years, the influence of human capital was confirmed by many economic studies, more precisely influence of education on economic growth and productivity. On the contrary, some experts criticize this unambiguous and inconclusive relation and some of them even disprove this hypothesis, publishing their studies. This paper considers the relationship of human capital and economic growth in Austria. The importance of human capital in the economic growth of this country is being proven by using two types of econometric analyses. One of them is cointegration test, which is followed by error correction model. In terms of concept simplification, the value of human capital is expressed by the level of education. We found out that positive long-run relationship exists between human capital and economic growth, i.e. national competitiveness.

Keywords: economic growth, human capital, cointegration, error correction model.

JEL Classification: J24, C30
AMS Classification: 62P20

1 Introduction

„Changing economic and social conditions have given knowledge and skills – human capital – an increasingly central role in the economic success of nations and individuals. … The key role of competence and knowledge in stimulating economic growth has been widely recognised by economists and others.“ [19]

The influence of education, actually of knowledge and skills respectively (human capital) to increase the productivity of individuals and the nation's wealth has been emphasized already by classical economists. For example, A. Smith in his book *An Inquiry Into the Nature and Causes of the Wealth of Nations* (1776) expressed the view that education is a form of investment that should bring individual returns which exceed the training costs and time spent learning [3]. Another important economist A. Marshall, in his book *Principles of Economics* (1890) also notes that the most valuable capital is the one which is invested in human beings [3].

Education, however, was considered in the past to be social consumption. The change in the view of education and comprehensive concept of human capital theory in particular, was brought by the Chicago School of economists especially, namely T. W. Schultz [22] with his book *Investment in Human Beings* (1962) and G. S. Becker in his book *Human Capital* (1964). Becker [8] defined human capital as skills and adequate motivation to apply these skills. The main premise, which the human capital theory is based on, postulates that education increases the productivity of the individual. Each individual tries to optimise the return on their investment in education and will continue in the study until the rate of return on her investment in education will exceed the rate of returns of alternative investments. The benefit to the society is then the increased labor productivity of better educated members of society and also technological progress. Schultz in his work focused primarily on the problems of developing countries, where he believes that the source of growth in these countries can be an investment in human capital. Generally, economists of the 1960's were trying to determine how various factors contribute to economic growth.

Human capital as a production factor is included in the new growth theory models. Total product in the endogenous growth models is determined by both physical capital and labor and human capital which is

[1] Silesian University in Opava, School of Business Administration in Karviná, Department of Economics, Univerzitní nám. 1934/3, 733 40 Karviná, Czech Republic, chudarkova@opf.slu.cz.
[2] Silesian University in Opava, School of Business Administration in Karviná, Department of Economics, Univerzitní nám. 1934/3, 733 40 Karviná, Czech Republic, verner@opf.slu.cz.

accumulated in through education in every individual. The new growth theory applies the extended Solow's model and production function presented in the following form [6]:

$$Y_t = (A_t L_t)^{1-\alpha-\beta} K_t^{\alpha} H_t^{\beta} \tag{1}$$

where Y is the product, A shows the level of technology, L is the labour, K is the physical capital, H is the human capital, α and β determine the proportions of individual factors on the overall product.

P. M. Romer, R. E. Lucas (see [17]) and also N. G. Mankiw, D. Romer and D. Weil contributed significantly to the development of endogenous growth models. Overall, the research in this area confirms the existence of a relationship between the development of education and economic growth. One of the conclusions of endogenous growth models is that economic growth depends partly on the level of human capital. It assumes that human capital is the source of production of new ideas. It is true that the more developed economy, the stronger the relationship of education to the economic growth. While in less developed countries the primary task of starting economic growth nationwide is to ensure primary education, in the developed countries on the other hand is to drive further economic growth primarily on ensuring tertiary education. Romer [21] in his work addresses the issue of differences between the education and experience on the one hand, and technological progress on the other. The main source of economic growth is technological progress, in his opinion. Mankiw, Romer and Weil [18] in their work tried to eliminate shortcomings of the Solow's model by including the human capital expressed as an investment in education. Simplified representation of the value of human capital, respectively identification of human capital investment in education, with the achieved level of education or the number of students in various stages of study, is often a prerequisite in empirical studies examining the human capital at the macroeconomic level.

Barro [4] and others find a strong positive correlation between schooling enrollment and the subsequent growth rate gross domestic product (GDP) per capita. Barro [5] states that the growth of human capital expressed as an average length of education by one year corresponds to an increase of GDP growth by four percentage points a year. Bassanini and Scarpetta [7] states, that their results point to a positive and significant impact of human capital accumulation to output per capita growth. If the average length of study period is ten years, one additional year of study will increase production by six per cent. The existence of correlations between human capital, in this case the number of university graduates, and economic growth in their work was also confirmed by De la Fuente and Donénech [11]. Through that research the need for investment in human capital can be justified. However, there are views that refute or do not confirm the influence of human capital on economic growth. Bils and Klenow [9] in their study do not disprove any correlation between economic growth and human capital. However, they concluded that it is the level of gross domestic product, respectively its growth, leading to a higher level of human capital in the economy. Unlike previous studies on the causality of these variables this one is seen in the reverse order. Söderbom and Teal [24] came to the conclusion that human capital has a small, and not statistically significant effect, on the level of output.

The main aim of this paper is to investigate the relationship between human capital and economic growth. We test the hypothesis whether educational expansion could affect economic growth in case of Austrian economy. Furthermore, we assume the higher economic growth the greater improvement in the competitiveness at the national level. There are many definitions and concepts of national competitiveness and therefore there is not only one approach and definition is still not unique [28]. In accordance with Porter [20] the only meaningful concept of macroeconomic competitiveness is national productivity. Based on this conception and relating approaches (e.g. Schwab [23]) someone could consider productivity and competitiveness as a synonym words. We can thus apply the approach in Hančlová [13] and express competitiveness easier way by means of GDP per capita.

The paper is structured as follows. In Section 2, we introduce and describe the dataset and specify the used methods. In Section 3, we present and discuss the results obtained from testing. Section 4 concludes the paper with summary of crucial findings.

2 Data and econometric methodology

We employ annual data for Austria between 1971 and 2008. GDP per capita at constant prices (2005) in USD was used as a proxy variable of economic growth, while for human capital we used tertiary education[3], more precisely enrolment in tertiary education and public expenditure per pupil in tertiary education (as a % of GDP

[3] There are two stages of tertiary education in ISCED 1997; the first contains tertiary programmes with academic orientation or occupation (ISCED 5) and the second contains tertiary studies that lead to a advanced research qualification – doctorate (ISCED 6). For more details see [27].

per capita). GDP at constant prices were collected from The United Nations database [25], population and education proxies from UNESCO database [26]. We converted all series into logs.

To examine the above mentioned relationship we perform cointegration test followed by error correction model (ECM)[4]. Similar to Asteriou and Agiomirgianakis [2], who examined the relationship in Greek economy. If variables are cointegrated, it indicates there is a long-run relationship or equilibrium. The ECM corrects for disequilibrium in short-run Gujarati [12]. If two or more series are integrated of order 1, i.e. I(1) and a linear combination of them is integrated of order 0, i.e. I(0), then the series are said to be cointegrated. I(1) means the series is non-stationary in level (has a unit root) but after first differencing, I(0) series is stationary in level [1].

To find out whether time series have unit root, i.e. are non-stationary we used augmented Dickey-Fuller test (ADF test), which estimates the following regression [12]:

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^{m} \alpha_i \Delta Y_{t-i} + \varepsilon_t \tag{2}$$

where ε_t is a white noise error term and where $\Delta Y_t = Y_t - Y_{t-1}$, etc. The null hypothesis tests whether $\delta = 0$.

Thus, we can perform cointegration test. Most specifically, we use the Johansen cointegration method (see [14, 15]). First of all we need to determine the lag order of vector autoregression through Akaike information criterion defined as [12]:

$$AIC = e^{2k/n} \frac{\sum \hat{u}_i^2}{n} = e^{2k/n} \frac{RSS}{n} \tag{3}$$

where k is the number of explanatory variables (including the intercept), n is the number of observations, RSS means the residual sum of squares and $2k/n$ the penalty factor.

The next step in Johansen`s approach is to estimate the VECM by maximum likelihood and the number of cointegrating vectors r. We applied two tests: (i) the trace test [10]:

$$\lambda_{trace}(r) = -n \sum_{i=r+1}^{m} \ln\left(1 - \hat{\lambda}_i\right) \tag{4}$$

and (ii) the maximum eigenvalue test:

$$\lambda_{\max}(r) = -n \ln\left(1 - \hat{\lambda}_{r+1}\right) \tag{5}$$

where n is the number of observations and $\hat{\lambda}$ is the i:th canonical correlation. The trace test tests the null hypothesis of at most r cointegrating vectors against the alternative more than r. The maximum eigenvalue tests the null hypothesis of r cointegrating vectors against the alternative of $r+1$. Tests reject the null hypothesis if $\lambda_{trace}(r)$ or $\lambda_{max}(r)$ are larger than their critical values or significance level.

As the next step the error[5] should be estimated from the cointegration equations (6) and test for unit root. The error need to be I(0) to confirm the long-run relationship.

$$e = Y_t - \alpha - \beta X_t \tag{6}$$

where e is error, α, β are long-run coefficients.

Finally, the ECM should be estimated [16].

$$\Delta Y_t = \varphi + \lambda_{e_{t-1}} + \omega_0 \Delta X_t + \varepsilon_t \tag{7}$$

where current changes in Y are a function of current changes in X and the degree to which the series are outside of their equilibrium in the previous period. The symbol e denotes the error from equation (6) and it can be

[4] Often also used – vector error correction model (VECM).
[5] Sometimes also called error correction term (ECT).

thought of as an equilibrium error. The model is out of equilibrium, if it is non-zero. The absolute value of λ shows us the correction to equilibrium in next period. Briefly, consider the case $\Delta X_t = 0$, $e_{t-1} < 0$ it indicates the dependent variable is below the equilibrium state, therefore λe_{t-1} should be positive and causes increasing Y in next period.

3 Empirical results

In this section we present and discuss the results from cointegration and ECM in the case of Austria and then compare and evaluate the development of used time series of Austria and the Czech Republic.

3.1 Austria

The ADF test of all time series implies that they are integrated of order one, I(1) at 5 % significance level. Akaike information criterion (AIC) indicates the optimal lag length of one period, i.e. one year[6]. Performing trace test and maximum eigenwalue test to estimate the VECM we yield one cointegration equation by trace test. Results of cointegration rank are shown in Table 1. The null and alternative hypotheses were established: H_0: r = 0 against H_1: r > 0. We reject the null hypothesis at 5 % significance level, therefore the number of cointegrating equations is not equal to zero. In next round we established: H_0: r = 1 against H_1: r > 1. We accepted null hypothesis, thus the number of cointegrating equation equals to one.

H_0	Trace statistic	5 % critical value	Probability 5 %
r = 0	35.2486	35.1928	0.0493
r = 1	19.0088	20.2614	0.0736

Table 1 Unrestricted cointegration rank trace test

Based on trace test, the long-run cointegration vector for our variables is given by following equation:

$$GDPPC_t = 0.903793 ENRTER_t + 0.581762 EXPTER_t - 3.083162 + e_t \quad (8)$$

and test error for unit root. ADF implies error is integrated of order 0, I(0), hence time series are cointegrated and long-run relationship confirmed[7]. Therefore we find a positive long-run relationship between economic growth and human capital and this conclusion is consistent with economic theory. Increasing enrolment in tertiary education (ENRTER) about 1 % will cause increasing GDP per capita (GDPPC) about 0.9 % and increasing about 1 % the public expenditure per pupil in tertiary education (EXPTER) will cause increasing GDP per capita almost 0.6 %. This conclusion confirms the hypothesis established in introduction.

ECM is given by equation (9):

$$\Delta GDPPC_t = 0.0444 e_{t-1} + 0.1864 \Delta GDPPC_{t-1} + 0.1691 \Delta ENRTER_{t-1} + 0.0678 \Delta EXPTER_{t-1} + \varepsilon_t \quad (9)$$

thus, GDP per capita is lower in short-run than in long-run. About 4.4 % of deviation from equilibrium is corrected next period by changes in GDP per capita.

3.2 Comparison with the Czech Republic

In this subsection we present the results of comparison of selected indicators in the Czech Republic and Austria. For the comparison we employ annual data during the period 1998 – 2008, respectively 1999 – 2008 in case of public expenditure[8]. Unfortunately, these data did not allow us to use the same method of evaluation as in the case of Austria. And so we compare the growth rates of economic growth and human capital. Development of these indicators is captured in Figure 1, Figure 2 and Figure 3. Based on these data, we can not confirm the relationship between human capital and economic growth in the Czech Republic unequivocally. A negative value of growth rate of enrolment in Austria was caused by the introduction of tuition fee in 2001.

[6] The same lag length chose [2].
[7] We conducted all needed tests but results are not presented for economy of space.
[8] We used this reduced time series because of former data for Czech Republic were not available.

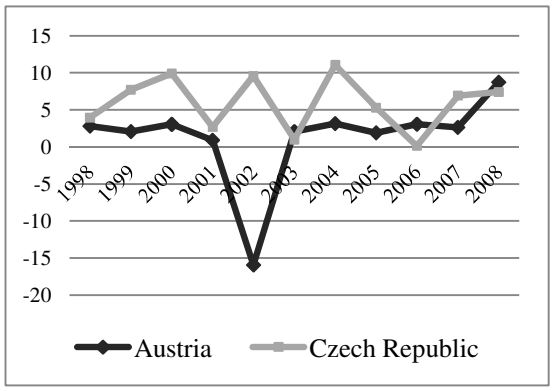


Figure 1 Growth rate of enrolment in total tertiary education (in %)

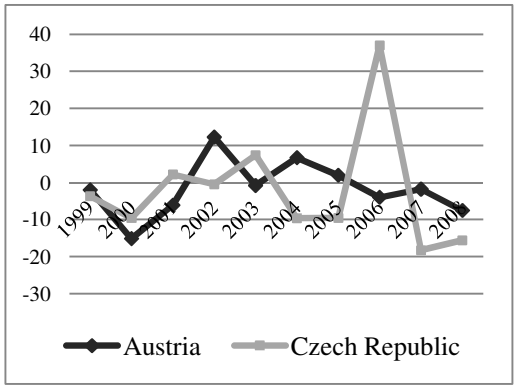


Figure 2 Growth rate of public expenditure on tertiary education per pupil as a % of GDP per capita (in %)

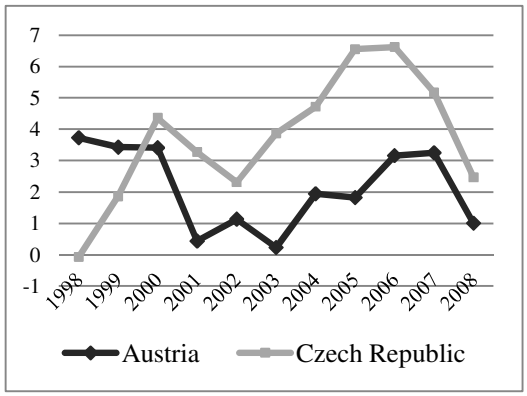


Figure 3 Growth rate of GDP per capita (in %)

4 Conclusion

The aim of this paper was to investigate the relationship between human capital and economic growth. We performed cointegration followed by error correction model. We employed annual data of Austrian economy (1971 – 2008) – GDP per capita for economic growth, enrolment in tertiary education and public expenditure in tertiary education as proxies for human capital. We found out, positive long-run relationship exists among these variables. Therefore, our hypothesis was confirmed. Moreover, according to our assumption, there is a positive relationship between human capital and national competitiveness. In the case of comparison with the Czech Republic we used the reduced time series (1998 – 2008) and based on these data, we can not confirm the relationship between human capital and economic growth in the Czech Republic unequivocally.

Acknowledgements

This paper was created in the students´ grant project "Seeking Factors and Barriers of the Competitiveness by Using of Selected Quantitative Methods". Project registration number is SGS/1/2012.

References

[1] Arlt, J.: Kointegrace v jednorovnicových modelech. *Politická ekonomie* **45** (1997), 733-746.
[2] Asteriou, D., and Agiomirgianakis, G. M.: Human Capital and Economic Growth. Time Series Evidence from Greece. *Journal of Policy Modeling* **23** (2001), 481-489.
[3] Balcar, J.: Lidský kapitál a jeho evoluce v ekonomické teorii. *Ekonomická revue* **9**, No. 3 (2006), 86-98.
[4] Barro, R. J.: *Economic Growth in a Cross Section of Countries*. NBER Working Paper No. 3120, 1989.
[5] Barro, R. J: Human Capital and Growth. *The American Economic Review* **91**, No. 2 (2001), 12-17.
[6] Barták, M., and Vomáčková, H.: *Ekonomický rozměr lidského kapitálu*. Universita J. E. Purkyně v Ústí nad Labem, Ústí nad Labem, 2007.
[7] Bassanini, A., and Scarpetta, S.: *Does Human Capital Matter for Growth in OECD Countries?: Evidence from Pooled Mean-Group Estimates*. OECD Economics Department Working Papers No. 282, 2001.
[8] Becker, G. S.: *Human Capital*. Columbia University Press, New York, 1964.
[9] Bils, M., and Klenow, P. J.: *Does Schooling Cause Growth or the Other Way Around?* NBER Working Paper No. 6393, 1998.
[10] Cipra, T.: *Finanční ekonometrie*. Ekopress, Praha, 2008.
[11] De la Fuente, A., and Donénech, R.: *Human Capital in Growth Regression: How much Difference Does Data Quality Make?* OECD Economics Department Working Papers No. 262, 2000.
[12] Gujarati, D.: *Basic Econometrics*. The McGraw-Hill Companies, New York, 2004.
[13] Hančlová, J., Kubicová, I., Macháček, M., Melecký, A., Melecký, L., Melecký, M., Nevima, J., and Ramík, J.: *Makroekonometrické modelování české ekonomiky a vybraných ekonomik Evropské unie*. VŠB-TUO, Ostrava, 2010.
[14] Johansen, S.: Statistical Analysis of Cointegrating Vectors. *Journal of Economic Dynamics and Control* **12** (1988), 231-254.
[15] Johansen, S.: Estimation and Hypothesis Testing of Cointegration Vectors in Gaussion Vector Autoregressive Models. *Econometrica* **59**, No. 6 (1991), 1551-1580.
[16] Koop, G.: *Analysis of Economic Data*. John Wiley & Sons, Chichester, 2005.
[17] Lucas, R. E.: On the Mechanics of Economic Development. *Journal of Monetary Economics* **22** (1988), 3-42.
[18] Mankiw, N. G., Romer, D., and Weil, D. N.: *A Contribution to the Empirics of Economic Growth*. NBER Working Paper No. 3541, 1990.
[19] OECD. *The Well-being of Nations: The Role of Human and Social Capital*. [Online]. [20th April 2012]. Available on: http://www.oecd.org/dataoecd/36/40/33703702.pdf.
[20] Porter, M. E.: *The Competitive Advantage of Nations: With a New Introduction*. Free Press, New York, 1998.
[21] Romer, P. M.: Endogenous Technological Change. *Journal of Political Economy* **98**, No. 5 (1990), S71-102.
[22] Schultz, T. W.: *Investment in Human Beings*. University of Chicago Press, Chicago, 1962.
[23] Schwab, K., et al.: *The Global Competitiveness Report 2010-2011*. World Economic Forum, Geneva, 2010.
[24] Söderbom, M., and Teal, F.: *Trade and Human Capital as Determinations of Growth*. CSAE Working Paper Series 2001-10, 2001.
[25] UNDATA. *United Nations Data Retrieval System*. [Online]. [9th May 2012]. Available on: http://data.un.org/Explorer.aspx
[26] UNESCO Institute for Statistics. *Predefined Tables*. [Online]. [9th May 2012]. Available on: http://stats.uis.unesco.org/unesco/ReportFolders/ReportFolders.aspx?CS_referer=&CS_ChosenLang=en
[27] UNESCO. *International Standard Classification of Education*. [Online]. [9th May 2012]. Available on: http://www.uis.unesco.org/Library/Documents/isced97-en.pdf.
[28] Verner, T.: National Competitiveness and Expenditure on Education, Research and Development. *Journal of Competitiveness* **3**, No. 2 (2011), 3-10.

Modelling of economic phenomena and dependences for corporate sustainable performance

Zuzana Chvátalová[1], Jiří Hřebíček [2]

Abstract. The determination of corporate sustainability performance and its key performance indicators is very important for company top management. The paper introduces chosen applications of quantitative methods in Maple for modelling of economic phenomena and dependences in an evaluation of company economic performance. There are discussed two examples of the evaluation of the Economic Value Added indicator (EVA) with Maple in the paper.

Keywords: mathematical modelling, economic phenomena, mathematical programming, Maple, key performance indicators, corporate sustainability performance.

JEL Classification: C88
AMS Classification: 90C08

1 Introduction

Verification of economic theory and real manifestations of the current economic practice can currently be seen as a power impulse for companies on market in the whole spectrum of their activity. Determining key performance indicators of companies, methods for their collection, analysis and correction of their measurement allows the modification of their development. They are becoming a solid base for analysis ensuring corporate sustainability of companies at the global market [3], [4], [6], [11]. The goals of the paper is introducing chosen applications of quantitative methods in Maple for modelling of economic phenomena and dependences in an evaluation of company economic performance and discuss them in the examples of the evaluation of the Economic Value Added indicator (EVA) [14], [15].

Maple is powerful mathematical software [1] for the construction of methods for mathematical modelling in economics [17] and finance engineering [18], numerical or graphical evaluation of key performance indicators (KPIs) of economic, environmental, social and corporate governance (ESG) phenomena in nature. Its statistical analysis and modelling of economic phenomena with visualization enable a company top management (e.g. chief executive officer further CEO) to manage and control more efficiently the given company [8]. Therefore, Maple is used in the research project "Construction of Methods for Multifactor Assessment of Company Complex Performance in Selected Sectors" (No. P403/11/2085) of Czech Science Foundation solved by solved by the Faculty of Business and Management of the Brno University of Technology (FBM BUT) and the Faculty of Business and Economics of the Mendel University in Brno (FBE MENDELU). It supports research of mathematical modelling of corporate sustainability performance for better application of project results in economic practice and also in the teaching of economic modelling [5], [7].

The paper introduces some applications of quantitative methods in Maple for modelling of economic phenomena and dependences in an evaluation of economic performance of company and discusses them in the example of the evaluation of the indicator EVA [14], [15].

2 Measurement of economic performance by Maple

In the Czech Republic dominates the classical approach of the measurement of economic performance, based on monitoring of standard indicators as *Return on Equity* (ROE), *Return on Investment* (ROI), *Return on Assets* (ROA), *Return on Sales* (ROS), *Return on Capital Employed* (ROCE), *Earnings before Interest and Taxes* (EBIT), *Earnings before Interest, Taxes, Depreciation and Amortization* (EBITDA), *Earnings after Taxes* or *Net Operating Profit after Tax* (EAT or NOPAT), *Earnings Per Share* (EPS) etc [10].

There exists on line *Benchmarking diagnostic indicators of financial system INFA* [19], which was developed in the collaboration between the Ministry of Industry and Trade (MIT) and the University of Economics, Prague (UEP). The MIT provides a data and programming capacity, and the UEP supports the methodological frame-

[1] Brno University of Technology, Faculty of Business and Management, Department of Informatics Faculty of Business and Management, Kolejní 4, 612 00 Brno, chvatalova@fbm.vutbr.cz.
[2] Mendel University, Faculty of Business and Economics, Department of Informatics, Faculty of Business and Economics, Zemědělská 1, 61300 Brno, hrebicek@mendelu.cz.

work and analytical work [13]. The INFA system was upgraded in April 2012. This innovation lies in the introduction of the Classification of Economic Activities (CZ-NACE) [20] and treatment algorithms, calculations of the change in the statements of the Czech Statistical Office (CSO).

The CSO collects from companies quarterly following economic data [8]: *Values of assets, liabilities and components; Acquisition of total fixed assets; Total revenues; Turnover; Turnover (revenues) from own goods and services; Turnover (revenues) from merchandise refers to revenues from sales; Total expenses; After tax economic result; Production value; Book value added; Margin; Book value employee; Return on assets (ROA); Return on equity (ROE); Return on sales (ROS); Return on costs; Expense-to-revenue; Inventories turnover time; Environmental protection investment; Non-investment environmental expenditure; Environmental protection expenditure; Expenditure on R&D*. These data are monitored by the CSO questionnaire P 3-04, which covers the entire size range of businesses and trades (calculated for companies with 0-19 employees, selected enterprises with 20-49 employees and all enterprises with 50 or more employees). Status data and selected data are monitored by the CSO questionnaire P 6-04 across the board in companies with significant assets regardless of the number of employees.

Kocmanová and Doležalová proposed the *Key Performance Indicators* (KPIs) in [11] for the measurement of economic performance of company in relation to the environmental, social and governance (ESG) indicators [6]. They established these economic KPIs on the basis of the results of empirical research of the project No P403/11/2085 in the last year. The economic KPIs in Tab. 1 provide quantitative forms of a feedback which reflect the company results in the framework of its corporate strategy. These KPIs can also help to the company Chief executive officer (CEO) to plan and manage company economic priorities, in particular, when these indicators are focused on the core business strategy, by means of operational plans, which include corporate performance targets.

No	Indicator name	KPI	Measurement
	Profit	EBIT	Earnings before Interest and Taxes
		EBITDA	Earnings before Interest, Taxes, Depreciation and Amortization.
		EAT	Earnings after Taxes or Net profit after Taxes
		EPS	Earnings Per Share, Price Earnings / (P/E Ratio).
EC2	Cash Flow	Free Cash Flow	EBIT - Taxes + Amortization - expenditure on the acquisition of fixed assets + /-increases in working capital.
		Operating Cash Flow	All the cash flows arising from the main activity of the company, which is the subject of its business (the movement of stocks, receivables, obligations).
EC3	Revenues	Total revenues	Revenues from own goods and services + Revenues from sale of merchandise (goods for resale) + Revenues of fixed assets + Revenues from sale of materials + Revenues of securities.
EC4	Turnover size	Turnover size	Revenues from own goods and services + Revenues from sale of merchandise (goods for resale) + Revenues of securities
EC5	Profit margin	Profit margin	The difference between turnover (revenues) from sales of goods and expenses on merchandise sold (i.e. on goods sold in the same condition as received).
EC6	Indicators of economic performance	Return on Equity	ROE = EAT / Equity
		Return on Investment	ROI = EBIT / Total capital
		Return on Assets	ROA = EBIT / Assets
		Return on Sales	ROS = EAT / Revenues
		Return on Capital Employed	ROCE = EBIT / Equity + Long-term liabilities
EC7	EVA	Economic Value Added	EVA = (ROE – Cost of Equity) * Equity

Table 1 KPIs of economic performance

The last indicator EVA in Tab. 1 presents the analysis of the *Economic Value Added* [13], [14], [15] an advanced evaluation method that measures the performance and the profitability of the company, taking in account the cost of capital that the company employs. This method, invented by Stern Stewart & Co. [21] is used today by more and more companies as a framework for their financial management and their incentive compensation system for the managers and the employees.

We developed Maple program [6], [16], where the indicator EVA is calculated with using indicators in Tab. 1 and data collected by the CSO. We issued from the following formula:

$$Economic\ Value\ Added = (Return\ on\ Equity - Cost\ of\ Entity) * Equity,$$

which can be reformulated to:

$$EVA = EAT - C * WACC = EBIT * (1 - t) - C * WACC,$$

where C denotes *Capital Employed,* t is *Corporate tax rate* and

$$EAT = NOPAT = Operating\ Profit * (1 - Tax\ Rate)$$

The parameter *WACC* (*Weighted Average Cost of Capital*) is calculated by:

$$WACC = (r_d * (1 - t) * D + r_e * E) / C,$$

where D denotes *Total debt and leases*, E is *Total market value of equity and equity equivalents or market capitalization*, $C = E + D$, r_d is *Required or expected rate of return on borrowings before taxes (cost of debt)* and r_e denotes *Cost of equity*.

We will use in Maple the *Capital Asset Pricing Model* (CAPM) [22] to calculate the *Cost of equity*. This model can be considered as a special case of *Markowitz Portfolio Model* [12]:

$$r_e = r_f + \beta\,(r_m - r_f)\,,$$

where denotes r_f – *Risk free rate*, r_m – *Expected market rate of return*, $(r_m - r_f)$ – *Market risk premium*, and β is a coefficient that measures the part of the asset's statistical variance that cannot be mitigated by the diversification (beta of a stock is or portfolio is a number describing the relation of its returns with that of the market as a whole).

We compared our above approach for the modelling of EVA indicator with EVA indicator calculated by the system INFA of the MTI (abbreviated EVA MIT) for a chosen anonymous company A. The complete description of Maple calculations and their visualization was made in Vecheta's master thesis [16]. The chosen results of this comparison are in Table 2.

year	2002	2003	2004	2005	2006	2007	2008
EVA MIT	11 673	11 673	11 673	11 673	11 673	11 673	11 673
EVA	11 673	7 217	7 880	4 577	24 117	10 530	-3 462
Relative difference	6%	53%	6%	20%	32%	6%	21%

Table 2 Comparison of EVA and EVA MIT indicators for company A [16]

The difference between both methods of calculations of the EVA indicator is caused by the different approach while calculating the *Costs of equity* because the CAPM method is based on subjective view of the analyst estimation apart from MIT method [13].

2.1 Statistical support of Maple for EVA

The *Statistics package* of Maple is a collection of tools for mathematical statistics, data analysis and statistical diagnosis [2]. Statistical computations in Maple combine the ease of working in a high-level, interactive environment with a very large and powerful set of algorithms. Large data sets can be handled efficiently with 35 built-in statistical distributions, sampling, estimations, data smoothing, hypothesis testing, and visualization algorithms. In addition, integration with the Maple symbolic engine means that you can easily specify custom distributions by combining existing distributions or simply by giving a formula for the probability or cumulative distribution function. The Statistics package also includes the *Data Analysis Assistant*, a graphical interface to the data analysis tools in this. It provides various commands for fitting models to data points and performing regression analysis based on least - squares methods.

We used *Curve Fitting Assistant* of Maple for least square fitting of EVA data and its visualization. This enabled us to analyze EVA trends of chosen company A from 2002 to 2008 year, Fig. 1.

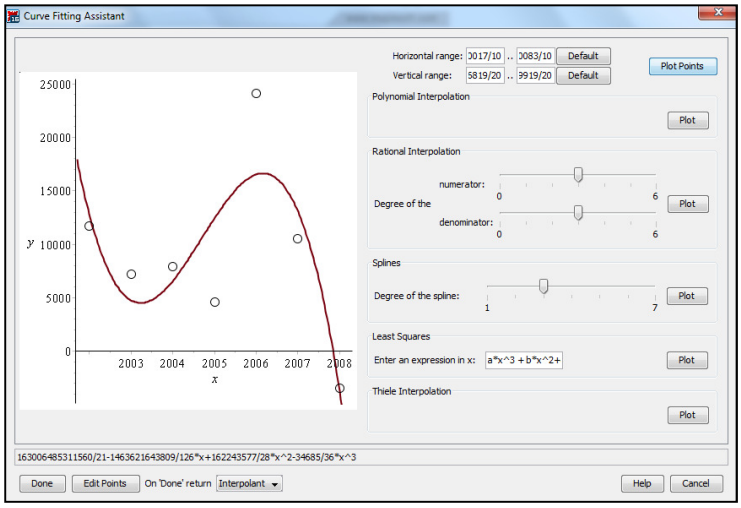


Figure 1 Fitting EVA data by least squares method

2.2 Use of neural networks in Maple at modelling EVA

The EVA indicator as a measure of surplus value created on an investment is based on the idea that a business has to cover both, the operating costs and the costs of capital. It stems from the estimate of economic, not the accounting profit.

We modified Maple program *A Feedforward Neural Network Forecasting Exercise from Application Centre* [23] from Maplesoft Application Centre based on methods published by Kendrick, Mercado and Amman [9] and applied this to our Maple program for calculation of EVA indicators [1], [5].

We divided input data of the given anonymous company B for our Maple program into three groups (teaching, testing, verifying). The output of Maple program was the value computed by using of neural network, which process was copying the process of computing of values of EVA indicator. The basic architecture of the neural network consists of a six layered perceptron working in the area of real numbers with linear rating of neurons. The neural network was created in Maple and it was composed of three coherent modules. The neural network used for storage of necessary data and intermediate calculations vectors and matrix. The matrix and vector character of data enables a simple using of cycles. At learning of the network for EVA indicator was made as a whole 20 223 study epochs, [1], [5].

Years	2002	2003	2004	2005	2006	2007	2008	2009
EVA MIT	2072	5749	14415	24179	32092	36658	22761	42265
EVA neural network	1927	7548	12672	22147	34445	43689	18961	36314
Relative Difference	7 %	31 %	12 %	8 %	7 %	19 %	16 %	14 %

Table 3 The comparison of EVA MIT and EVA neural network indicators for company B

3 Conclusion

The measurement of company performance is very important for company top management. The combination of suitable information technology, the use of economic and quantitative methods and the correct interpretation of results can improve the CEO decision making processes. Application of scientific quantitative methods in business companies has interdisciplinary impact. Modern enterprise information systems contain large amount of data and information that can be used more efficiently thanks to the use of Maple. We introduced this on the calculation the indicator EVA in two examples. We can see that the sophisticated financial analysis of the company and the comparison with its competitors in Maple is valuable source of information for company managers.

Acknowledgements

This paper is supported by the Czech Science Foundation. Name of the Project: Construction of Methods for Multifactor Assessment of Company Complex Performance in Selected Sectors. Reg. No. P403/11/2085.

References

[1] Bartulec, T.: *Mathematical Modeling of Company Efficienty Using Neural Networks in Maple*. (master's thesis in Czech), FBM BUT, Brno, 2011.

[2] Bernardin, L., Chin, P., DeMarco, P., Geddes, K. O., Hare, D. E. G., Heal, K.M., Labahn, G., May, J. P., McCarron, J., Monagan, M. B., Ohashi, D., and Vorkoetter, S. M.: *Maple 16 Programming Guide*.Maplesoft, Waterloo, 2012.

[3] Griffiths, W. E.: *Learning and Practicing Econometrics*. John Wiley & Sons Inc., New York, 1993.

[4] Hey, J. D.: *Intermediate Microeconomics: People are Different*. Pearson Custom Publishing, York, 2011.

[5] Hřebíček, J., and Chvátalová, Z.: Chosen Features of Maple for Mathematical Modeling in Education and Practice. *Journal of Applied Mathematics* **4** (2011), 379-393.

[6] Hřebíček, J., Soukopová, J., Štencl, M., and 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.

[7] Chvátalová, Z., and Hřebíček, J.: Education of Economics with Maple. In: *Proceedings of the International Conference on Environment, Economics, Energy, Devices, Systems, Communications, Computers, Pure and Applied Mathematics* (Lazard M. Et al., eds.). WSEAS Press, Florence, 2011, 435-440.

[8] Chvátalová, Z., and Šimberová, I.: Economic phenomena via mathematical modelling in Maple system, *Business, Management and Education* **9** (2011), 260–276.

[9] Kendrick, D. A., Mercado, P. R., and Amman, H. A.: *Computational Economics*. Princeton University Press, Princeton, 2006.

[10] Kocmanová, A., Hornugová, J., and Klímková, M.: *Sustainability:The Integration Environmental, Social and Economic Performance of Company* (in Czech). CERM, Brno, 2010.

[11] 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* **60** (2012) to appear

[12] Marinič, P.: *Plánování a tvorba hodnoty firmy*. Grada, Praha, 2008.

[13] Neumaierová, I., and Neumaier, I.: From benchmarking diagnostic system of financial indicators INFA to the dynamic scorecard INFA (in Czech). In: *Proceedings Finance a výkonnosti firem ve vědě, výuce a praxi*. (Jirčíková, E., ed.). Univerzita Tomáše Bati, Zlín, 2007.

[14] Rogerson, W. P.: Intertemporal Cost Allocation and Managerial Investment Incentives: A Theory Explaining the Use of Economic Value Added as a Performance Measure. *Journal of Political Economy* **105** (1997), 770-795.

[15] Růčková, P.: *Financial analysis: methods, indicators, used in practice* (in Czech). 3rd update ed., Grada Publishing, Prague, 2009.

[16] Vecheta, L.: *Finance analysis of company with Maple* (master's thesis in Czech). FBM BUT, Brno, 2010.

[17] Application centre Maplesoft: Economics [online]. [cit. 2012-05-30]. Available at http://www.maplesoft.com/applications/Category.aspx?cid=218

[18] Application centre Maplesoft: Financial Engineering [online]. [cit. 2012-05-30]. Available at http://www.maplesoft.com/applications/Category.aspx?cid=219

[19] http://www.mpo.cz/cz/infa.html

[20] http://www.czso.cz/csu/klasifik.nsf/i/klasifikace_ekonomickych_cinnosti_%28cz_nace%29

[21] http://en.wikipedia.org/wiki/Capital_asset_pricing_model

[22] http://www.maplesoft.com/applications/view.aspx?SID=32976&view=html

Estimation of weights in multi-criteria decision-making optimization models

Vladislav Chýna[1], Martina Kuncová[2], Jana Sekničková[3]

Abstract. Multi-criteria evaluation of alternatives is a category of the operational research that uses various kinds of methods to find the best alternative, the order of the alternatives or divide them into efficient and inefficient according to the selected criteria. A lot of methods need cardinal information about the criteria – it means weights of the criteria – to calculate the results. The weight vector describes the importance of criteria and its influence over the results may be crucial. It is easy to find non-dominated alternatives but it is not easy to say which of them can be on the top. Sometimes it is good to know whether there exists a weight vector for the selected alternative to be at the first place. That is why we have decided to think about the method that tells us what weight vector should be used for the selected alternative to be on the first place. This article describes the optimization models for this situation when selected methods of multi-criteria evaluation of alternatives are used.

Keywords: Multi-criteria Evaluation of Alternatives, Monte Carlo Simulation, Mobile Phone Tariffs

JEL Classification: C44, C15
AMS Classification: 91B06, 65C05

1 Introduction

In economy we must face a lot of decisions that have to be made, and pay a lot of money afterwards often without knowing whether we have done right or wrong. When everything is given, the solution or decision can be based on the common sense or on the solution of some mathematical model. But the problem is that a lot of things not only in economy are not certain – especially when we think about money spent for phoning. People are usually able to describe the length of their calls "something between 50 and 300 minutes per month" or "150 minutes at a medium". Although it seems to be vague, inaccurate and insufficient, with some knowledge of statistical distributions we are able to use given information and even make a decision or recommendation via Monte Carlo simulation model. On the other hand there are some methods for decision-making that also can help in this situation – and here the preferences must be specified for example by the weights of the criteria. But is it possible to use multi-criteria evaluation of alternative methods to obtain the same results as from the simulation model?

Simulation modeling and multi-criteria evaluation of alternatives are two different principles of mathematical methods connected with the operational research. Monte Carlo simulation tries to iteratively evaluate the deterministic model by using random inputs. Methods of multi-criteria evaluation of alternatives use given inputs to find the best alternative or the order of the alternatives with respect to the given criteria and weights. In this article we try to find the optimal mobile phone tariff for the given employee by Monte Carlo simulation and also by selected multi-criteria evaluation of alternatives methods. As weights are necessary we solve an optimization model to find the weight vector for the selected tariff to be on the first place. The main question is if it is possible to use these different principles to find the same results.

We will describe the simulation model and we compare the results obtained from simulation in MS Excel and Crystal Ball with the results taken from the static decision-making model when the WSA and TOPSIS methods are used and when the optimization model is created to find the right weight vector for the selected alternative to be the best.

[1] University of Economics Prague, Dpt. of Econometrics, W.Churchill Sq. 4, 13067 Prague 3, vladislav.chyna@vse.cz
[2] University of Economics Prague, Dpt. of Econometrics, W.Churchill Sq. 4, 13067 Prague 3, martina.kuncova@vse.cz
[3] University of Economics Prague, Dpt. of Econometrics, W.Churchill Sq. 4, 13067 Prague 3, jana.seknickova@vse.cz

2 Methods and data

Before we start the analysis we have to select the alternatives (mobile operators' tariffs), the criteria and the distributions for the random variables generation. Our analysis is aimed at the specific situation – to find the best tariff for one employee of the Executive Board of the Czech Union for Nature Conservation to minimize the costs of telephone calls. The entire model for more employees has been created in the diploma thesis [5] where all (69 possible) the mobile operators' tariffs and their data are described. We have selected the employee whose calls are somewhere between 20 and 1200 minutes per month (usually no SMS).

The problem occurs in the case when we don't know preferences of user in any form. Also in such case one solution of this problem is a simulation of weights as we have tried in [7]. The random generation is one possibility, the other one is to use an optimization model that calculates the utility of each alternative (by WSA) or the distance from the ideal alternative (TOPSIS).

2.1 Monte Carlo Simulation

Simulation methods belong to the suitable instruments that can be used in the real world situations to better understand the reality or to make a responsible decision. Simulation nowadays means 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 reasons for this are various: 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. Simulation is used in many contexts, including the modeling of natural or human systems in order to gain insight into their functioning (manufacturing, automobile industry, logistics, military, healthcare, etc.), simulation of technology for performance optimization, safety engineering, testing, training and education.

The problem of some economic models is the lack of the information – especially in the retail sector sometimes only managers themselves know how the process works, what the typical number of customers during a period is etc. In this kind of situations we cannot use basic statistical or mathematical models as we do not have the strict or real data. That is why Monte Carlo simulation can help as it uses random variables from different distributions. Monte Carlo simulation (or technique) is closed to statistics as it is a repeated process of random sampling from the selected probability distributions that represent the real-life processes [8]. On the basis of the existed information we should select the type of probability distribution that corresponds to our expectations and define all the parameters for.

The usage of MS Excel and Crystal Ball for the mobile phone tariffs is described in [6]. This kind of simulation was used also in the diploma work [5] to find the best tariff. But it is possible to use it also to generate the weights of the criteria – or better to say generate the points for each criterion and then calculate the weights using the Point method [3].

2.2 Multi-criteria evaluation of alternatives

Multi-criteria evaluation of alternatives belongs to the category of discrete multi-criteria decision making models where all the alternatives (a_1, a_2, …, a_p) and criteria (f_1, f_2, …, 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 [2], [3], [4], the two that we use are WSA and TOPSIS.

WSA (Weighted Sum Approach)

One particular example of utility maximization methods is called WSA and it is based on assumptions of linearity and maximization of all the partial utility functions. Therefore the minimizing criteria need to be transformed into maximizing criteria. Then the decision matrix $\mathbf{Y} = (y_{ij})$ is transformed into a normalized decision matrix $\mathbf{R} = (r_{ij})$, in which all the elements use the same units of measurement:

$$r_{ij} = \frac{y_{ij} - D_j}{H_j - D_j}, \quad r_{ij} \in \langle 0;1 \rangle, \quad \forall\, i = 1, \ldots, p, \quad j = 1, \ldots, k,$$

where r_{ij} denotes normalized value for the i-th alternative and j-th criterion, D_j – basal value, the worst possible value an alternative acquires in the j-th criterion, and H_j – ideal value, the best possible value an alternative acquires in the j-th criterion. Obviously, $r_{ij} = 0$ for the basal alternative, and $r_{ij} = 1$ for the ideal alternative.

The next step consists in calculation of the utility that can be cumulated from each alternative using the formula:

$$u(a_i) = \sum_{j=1}^{k} v_j \cdot r_{ij}, \quad \forall\, i = 1, \ldots, p,$$

where v_j denotes corresponding element from the weight vector and r_{ij} denotes normalized value gained from previous step. Obviously, the alternative with the highest value of utility is considered compromise. In addition, WSA makes it possible to arrange all the alternatives with respect to their utility values.

TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution)

The output provided by TOPSIS is a complete arrangement of possible alternatives with respect to the distance to both the ideal and the basal alternatives incorporating relative weights of criterion importance. The required input information includes decision matrix **Y** and weight vector **v**. In addition, in the same way as in the WSA an assumption of maximization of all the criteria is true (otherwise it is necessary to make an appropriate transformation). This decision-making approach can be summarized in the following steps (detailed description of steps and notation in [7]:

- normalize the decision matrix according to Euclidean metric:

$$r_{ij} = \frac{y_{ij}}{\sqrt{\sum_{i=1}^{p} y_{ij}^2}}, \quad \forall\, i = 1, \ldots, p, \quad j = 1, \ldots, k,$$

- calculate the weighted decision matrix $\mathbf{W} = (w_{ij}) = v_j \cdot r_{ij}$, and from the weighted decision matrix **W** identify vectors of the hypothetical ideal **H** and basal **D** alternatives over each criterion

- measure the Euclidean distance of every alternative to the ideal and to the basal alternatives over each attribute:

$$d_i^+ = \sqrt{\sum_{j=1}^{n} (w_{ij} - H_j)^2} \quad and \quad d_i^- = \sqrt{\sum_{j=1}^{n} (w_{ij} - D_j)^2}, \quad \forall\, i = 1, \ldots, p,$$

- for all alternatives determine the relative ratio of its distance to the basal alternative:

$$c_i = \frac{d_i^-}{d_i^+ + d_i^-}, \quad \forall\, i = 1, \ldots, p,$$

- rank order alternatives by maximizing ratio c_i .

2.3 Setting of weights for the winner

As we mentioned in the introduction the preferences of decision maker can be modeled by weight vector. In this part of this paper we would like to find a weight vector for the selected alternative to be the best. Let assume we know list of alternatives (a_1, a_2, …, a_p), list of criteria (f_1, f_2, …, f_k), decision matrix Y and also the winning alternative a_q. Note that this alternative have not to be a winner, we only wish it will be winner. In both method WSA and TOPSIS we can immediately transform minimizing criteria into maximizing and then normalize the decision matrix according to previous description. For both steps weight vector is unknown.

The WSA model for setting of weight vector

In this model v_j denotes weights and they are the variables of this model. The optimization models differ according to aim of optimization. We can search the weights that are enough large for all aternative (maximize ε) or we can search such weights that the winner has maximal utility difference D (difference between winner and the alternative on the second place). The model has a following form:

$$\max \varepsilon \quad or \quad \max D$$
$$subject\ to$$
$$\sum_{j=1}^{k} v_j = 1,$$
$$u_i = \sum_{j=1}^{k} v_j \cdot r_{ij}, \quad \forall\, i = 1, \ldots, p,$$
$$u_q \geq u_i + D, \quad \forall\, i \neq q, i = 1, \ldots, p,$$
$$v_j \geq \varepsilon, \quad \forall\ j = 1, \ldots, k,$$
$$\varepsilon \geq 0, D \geq 0.$$

It is known the problem of non-universality of WSA method and so there can exist non-dominated solutions that cannot be the winners. In such case the output of this model will be „no feasible solution". This problem can be solved by using of different method, e.g. TOPSIS.

The TOPSIS model for setting of weight vector

In this model v_j again denotes weights and they are the variables of this model. We can also search the weights that are enough large for all aternative (maximize ε) or we can search such weights that the winner has maximal difference of the relative ratio of distance to the basal alternative (difference between winner and the alternative on the second place). The model has a following form:

$$\max \varepsilon \quad or \quad \max D$$
$$subject\ to$$
$$\sum_{j=1}^{k} v_j = 1,$$
$$w_i = \sum_{j=1}^{k} v_j \cdot r_{ij}, \quad \forall\, i = 1, \ldots, p,$$
$$d_i^+ = \sqrt{\sum_{j=1}^{n} (w_{ij} - H_j)^2}, \quad \forall\, i = 1, \ldots, p,$$
$$d_i^- = \sqrt{\sum_{j=1}^{n} (w_{ij} - D_j)^2}, \quad \forall\, i = 1, \ldots, p,$$
$$c_i = \frac{d_i^-}{d_i^+ + d_i^-}, \quad \forall\, i = 1, \ldots, p,$$
$$c_q \geq c_i + D, \quad \forall\, i \neq q, i = 1, \ldots, p,$$
$$v_j \geq \varepsilon, \quad \forall\ j = 1, \ldots, k,$$
$$\varepsilon \geq 0, D \geq 0.$$

The case of max ε has always the solution (weights can be zero) in comparison to max D. Unfortunatelly, this model is non-linear in contrast to WSA model. Therefore, to find the solution is not so easy as in WSA model. In the special case when the basal alternative is zero for all criteria we can use a binary model, where

$$\sum_{i=1}^{p} x_{ij} \geq 1, \quad \forall\ j = 1, \ldots, k,$$
$$w_{ij} \geq D_j - M \cdot (1 - x_{ij}), \quad \forall\, i = 1, \ldots, p, \quad \forall\ j = 1, \ldots, k,$$
$$w_{ij} \leq D_j, \quad \forall\, i = 1, \ldots, p, \quad \forall\ j = 1, \ldots, k,$$
$$x_{ij} \subset Bi = \{0; 1\},$$

and M is an enough large constant.

3 Results and Discussion

The first part of the analysis was the Monte Carlo simulation to find the best tariffs for the given situation. The second task was to create optimization models to finding weights that represent preferences of decision maker used in simulation model.

3.1 Monte Carlo Simulation Results

As we know the number of minutes called per month vary between 20 and 1200, we have used the generation of random variables from the uniform distribution with parameters (20; 1200) to obtain the number of minutes called. The probability of the calling to all operator's networks were given and are in Table 1 (as U:fon is the smallest operator, the given employee does not call to this network at all). According to them the minutes to each network have been divided, then free minutes have been subtracted, the rest has been multiplied by the price per minute and the monthly fee has been added to obtain the total price.

Calls to / TOTAL minutes	**1134.27487**	**percent**
O2	306.254214	0.27
T-mobile	317.596963	0.28
Vodafone	170.14123	0.15

Table 1 Example of the generated minutes and their distribution into call to the networks

The best tariffs are in the Table 2, the minimal cost per month for these tariffs are 1500-1800 CZK. The best tariffs from other operators start at 1900 CZK (O2, Vodafone).

Operator	**Tariff**	**Monthly fee (CZK)**	**Free minutes per month**	**Price per minute call (CZK)**
T-mobile	Podnikatel Plus 1100	1320	110 (own)	2.52
T-mobile	Podnikatel Plus 700	840	70 (own)	3
U:fon	Unifon Duo 2/2	0	20	2.9

Table 2 The best tariffs from Monte Carlo simulation (uniform distribution of minutes called per month)

We have also tried different distribution – triangular with parameters (20; 500; 1200) minutes (as we know the average length of monthly calls is 500 minutes). To obtain the results we have used Crystal Ball. The results are nearly the same as in the previous case – the comparison of the best tariffs is at the Figure 1. The average monthly costs are about 1600 CZK, but they can vary from 450 to 3000 CZK, so we cannot say the exact order of the tariffs, but only show these cheapest (because the other tariffs started from 1500 to 6000 CZK). From T-mobile the best ones are the tariffs "Podnikatel plus 450", "Podnikatel plus 700", "Podnikatel plus 1100" and "Kredit 1000", from Vodafone the tariff "400 minutes", from O2 the tariffs "Neon L+" and "Podnikatel XL", from U:fon the tariffs "Unifon" and "Unifon Duo 2/2".

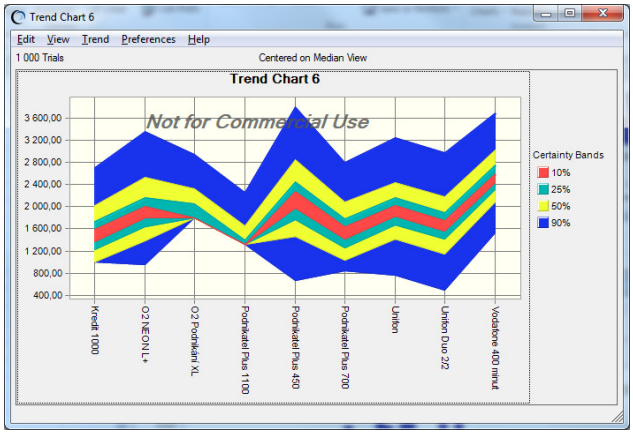


Figure 1 Comparison of the best tariffs from Crystal Ball

3.2 Optimization Model Results

According to subsection 3.1 we used models from subsection 2.3 for setting of weights that model preferences of user. The results for tariff Unifon Duo 2/2 are included in Table 3. Unfortunately, there exist no solution (no weights) for tariffs T-mobile Podnikatel Plus 1100 and 700 using WSA. It is the case mentioned in section 2.3.

The solution for TOPSIS non-binary model we try to search for a long time by using LINGO 12.0 with no success (the reason is given by non-linearity of the model). By using binary model we have obtained the results in Table 3.

weight	v_1	v_2	v_3	v_4	v_5	v_6
WSA (max ε)	0.7384	0.0490	0.0490	0.0490	0.0657	0.0490
WSA (max D)	0.4621	0.0116	0.0116	0.4613	0.1159	0.0419
TOPSIS (max ε)	0.4861	0.1028	0.1028	0.1028	0.1028	0.1028
average weight	0.4841	0.0545	0.0545	0.2044	0.0948	0.0646

Table 3 The results for Unifon Duo 2/2

From all results (using by WSA and TOPSIS) we can concluded that for our decision maker the first criterion (fixed payment tariff) has the highest weight (more than 46 %). The second criterion is the fourth one (the number of free minutes), the third place is reserved for the fifth criterion (advantages in own net) and the forth for the last criterion (advantages in other net). The last two criteria (price for 1 minute calling in own net and price for 1 minute calling in other net) have the same however lowest importance. All mentioned criteria are explicitly formulated in [7].

4 Conclusion

In this paper we have presented the fact that the preferences of given decision maker can be modeled by using optimization models. We simulate real situation and by Monte Carlo simulation we observe the best alternatives of mobile phone tariffs. By using optimization model we have found the weight vector that corresponds to decision maker preferences and the given alternative found by multi-criteria evaluation of alternative is the best for him or her.

Unfortunately, we have illustrated that such vector cannot exist in the case we use WSA method. In the case of TOPSIS such vectors exist but it is not easy to find them although using of optimization software due to non-linearity of model. But this methodology can be successfully used for modeling of decision maker preferences in the case that these preferences are unknown.

Acknowledgements

This work was supported by the Grant Agency of the Czech Republic GAČR No. 403/12/1387 and also by the Internal Grant Agency of the University of Economics Prague IGA F4/16/2011 (IG403011).

References

[1] Banks, J.: *Handbook of Simulation*. USA, John Willey & Sons, 1998.
[2] Evans, G.W.: An Overwiev of Techniques for Solving Multiobjective Mathematical Programs. *Management Science*. 1984, vol. 30, No. 11, 1268-1282.
[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] Harák, T.: *Využití simulací k redukci nákladů za telefonování v reálné firmě*. Diploma Thesis, University of Economics, Prague 2012
[6] 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. *Proceedings of Mathematical Methods in Economics 2006 Conference*, pp. 323-332. September 13-15, Pilsen (Czech Republic), 2006.
[7] Kuncová, M., Sekničková, J.: Multi-criteria Evaluation of Alternatives Applied to the Mobile Phone Tariffs in Comparison with Monte Carlo Simulation Results. *Proceedings of Quantitative Methods in Economy 2012 Conference* , May 30 – June 1, Bratislava (Slovak Republic), 2012
[8] Turban , E., Meredith, J., R.: *Fundamentals of Management science*. 6th ed. USA, Richard D.Irwin Inc., 1994.

Financial stability indicator predictability by support vector machines

Kristýna Ivanková[1]

Abstract. Support Vector Machines are a successful machine-learning algorithm used for classification, regression and prediction of time series. We optimize learning parameters and select the feature set with the smallest predictive error. We also explore the development of errors for predictions with larger time skips.

We've applied the method for the prediction of CISS (Composite Indicator of Systemic Stress), a stability indicator created by the European Central Bank. We've chosen this indicator among other state-of-the-art indicators because of its high frequency, which indicates a quick response to distinctive changes on the market.

The results show that CISS can be partially explained just by its past values up to six to eight weeks ahead, but large behaviour shifts are still surprising for the model. We've also discovered that including data over three months of age into the prediction context won't improve the results.

Keywords: SVM, prediction, financial stability, systemic stress, CISS.

JEL classification: C13, C45, G32.
AMS classification: 68T10, 62H30, 62P20.

1 Introduction

The ongoing financial crisis has motivated lots of financial instability research in the recent years. Many authors began to study the notions of systemic risk and systemic stress. Referencing the definitions from Holló [4]:

> Systemic risk can be defined as the risk that financial instability becomes so widespread that it impairs the functioning of a financial system to the point where economic growth and welfare suffer materially. Systemic stress is interpreted as that amount of systemic risk which has materialised.

Systemic stress is usually accompanied by increases in uncertainty, disagreement among investors, asymmetry of information between borrowers and lenders, risk aversion, flight to quality or flight to liquidity.

One of the recurring research themes is the introduction of novel indicators of financial stability (or financial stress). These indicators capture the current or past financial situation. Some examples of actively used indicators follow.

- The Laeven-Valencia index of systemic banking crises introduced in [7] is a broad, coincident indicator for systemic financial crises. It covers 169 countries and specifies the presence or absence of a crisis with yearly frequency since 1970.
- The Financial Stress Index (FSI) for 17 advanced economies [3] and its modified version for 26 emerging economies [1] introduced by the International Monetary Fund. Both are monthly indicators of national financial system strain and rely on price movements relative to past levels or trends. The FSI-AE index is defined for most of its countries since 1980, the FSI-EM since the end of 1996.

[1]Charles University in Prague, Institute of Economic Studies, Opletalova 26, 110 00, Praha 1, kristyna.ivankova@gmail.com

- A Composite Indicator of Systemic Stress (CISS) in the financial system [4] introduced in the European Central Bank. This is a short-term (weekly) composite systemic stress measure for the EU sector aggregating five market-segment subindices: money market, bond market, equity market, financial intermediaries and foreign exchange market. The combined index is based on cross-correlation of subindices after a rank transformation. The original CISS is defined only after the introduction of the Euro (1999), but the authors present a version extended up to 1987.

It would prove useful to be able to predict systemic stress and handle an emerging crisis before it actually breaks out. We will focus on CISS since the other indicators are too sparse for our purposes.

In section 2 we present a theoretical overview of our prediction methodology and in Section 3 we discuss its application to the prediction of CISS for various delays and present the obtained results.

2 Support vector machine regression

Support Vector Machines (SVMs) are a machine learning method introduced by Vapnik [10]. Good introductions into the SVMs are [2], [9]; for an application of SVMs to prediction of time series see [8].

Suppose we are given examples $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$, where $\mathbf{x}_i \in \mathbb{R}^d$ are the training contexts and $y_i \in \mathbb{R}$ are training targets. We seek a function $f(\mathbf{x})$ that maps contexts to targets and will predict new, previously unseen data well. Predictions $\tilde{y}$ for an unseen context $\tilde{\mathbf{x}}$ will be done by evaluating $f(\tilde{\mathbf{x}})$.

For the special problem of predicting a time series, the contexts $\mathbf{x}_i$ may consist of past values, their transformations, and external data.

2.1 The linear case

Let us first consider the linear case $f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b$ where $\langle \cdot, \cdot \rangle$ denotes the dot product. Following Smola and Schölkopf [9], we will penalize training errors by the standard ε-insensitive loss function ([9], Equation 4)

$$|z|_\varepsilon = \begin{cases} 0 & \text{if } |z| \leq \varepsilon, \\ |z| - \varepsilon & \text{otherwise.} \end{cases}$$

We prevent overfitting by adding the regularization term $\|w\|^2$. The trade-off between accuracy on the training set and model complexity will be controlled by the regularization constant $C > 0$.

We can now formulate the optimization problem ([9], Equation 3):

$$\begin{aligned} \underset{\mathbf{w},\, b,\, \xi_i,\, \xi_i^*}{\text{minimize}} \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ \text{subject to} \quad & -\varepsilon - \xi_i \leq \langle \mathbf{w}, \mathbf{x}_i \rangle + b - y_i \leq \varepsilon + \xi_i^*, \quad i \in \{1, \ldots, n\} \\ & \xi_i,\ \xi_i^* \geq 0, \quad i \in \{1, \ldots, n\}. \end{aligned}$$

The Lagrange dual problem has the form

$$\begin{aligned} \underset{\alpha_i,\, \alpha_i^*,\, \eta_i,\, \eta_i^*}{\text{maximize}} \quad \inf_{\mathbf{w},\, b,\, \xi_i,\, \xi_i^*} \Big[& \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ & - \sum_{i=1}^{n} \alpha_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b - y_i + \varepsilon + \xi_i\right) \\ & + \sum_{i=1}^{n} \alpha_i^* \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b - y_i - \varepsilon - \xi_i^*\right) \\ & - \sum_{i=1}^{n} (\eta_i \xi_i + \eta_i^* \xi_i^*) \Big] \\ \text{subject to} \quad & \alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0, \quad i \in \{1, \ldots, n\}. \end{aligned}$$

The gradient of the Lagrangian function with respect to the primary variables must vanish at the infimum. This gives the following set of constraints:

$$\begin{aligned} \sum_{i=1}^{n}(\alpha_i^* - \alpha_i) &= 0, \\ \mathbf{w} &= \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\mathbf{x}_i, \\ \eta_i &= C - \alpha_i, \\ \eta_i^* &= C - \alpha_i^*. \end{aligned}$$

After substitution we finally arrive at ([9], Equation 10):

$$\begin{aligned} \underset{\alpha_i,\,\alpha_i^*}{\text{maximize}} \quad & -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\,\langle \mathbf{x}_i, \mathbf{x}_j\rangle \;-\; \varepsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) \;+\; \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*) \\ \text{subject to} \quad & \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0, \\ & 0 \le \alpha_i \le C, \quad 0 \le \alpha_i^* \le C, \quad i \in \{1, \ldots, n\}. \end{aligned}$$

This is a quadratic convex programming problem with linear constraints and a dense dot product matrix and as such it can be solved using well-known quadratic programming algorithms. For notes about computing b see [9].

Let us note that $f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x}\rangle + b$ can be rewritten as $\sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\langle \mathbf{x}_i, \mathbf{x}\rangle + b$. Furthermore, it can be shown [9] that Lagrange multipliers α_i, α_i^* that correspond to input vectors lying inside the ε-insensitive region $|f(\mathbf{x}_i) - y_i| < \varepsilon$ vanish in the solution. This means that only a subset of the input examples is needed to describe f; examples from this subset are called the *Support Vectors*.

2.2 The non-linear case

We will now describe the *kernel trick* which makes SVMs the algorithm of choice for many machine-learning settings.

We want to create a (generally non-linear) mapping ϕ that maps inputs $\mathbf{x}_i$ into a high-dimensional *feature space* F, and perform the whole computation in F.

We've established earlier that both the optimization problem and function evaluation require only dot products between input examples. The trick is to choose a *kernel* function that corresponds to the dot product in F,

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j)\rangle_F,$$

and can be evaluated directly from the contexts without computing the map ϕ. This allows us to perform computations in the larger space F with no extra costs.

The most popular kernel choices for SVM regression are

- polynomial kernels $\quad k(\mathbf{x}_i, \mathbf{x}_j) = \big(\langle \mathbf{x}_i, \mathbf{x}_j\rangle + 1\big)^p$,
- Gaussian radial basis function (RBF) kernels $\quad k(\mathbf{x}_i, \mathbf{x}_j) = \mathrm{e}^{-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2}$, and
- sigmoidal kernels $\quad k(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\kappa\langle \mathbf{x}_i, \mathbf{x}_j\rangle - c)$.

For problems with a large number of training examples, the basic linear kernel $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j\rangle$ also enjoys some popularity because its special structure enables significant speed improvements (see e.g. [6]).

3 Application

We've divided the CISS dataset ***a*** into the training set (from Jan 1991 to Sep 2007) and the test set (Oct 2007 to Dec 2010). We've tried to learn a model for the CISS value δ weeks ahead:

$$y_i = a_{i+\delta} \quad \text{with } \delta \in \{1, 2, 4, 6, 8\}.$$

To find the SVM model with the least predictive error, we need to optimize the learning parameters. These consist of

- the margin width ε,
- the regularization parameter C, and
- the kernel parameters.

The objective function was chosen to be the average absolute prediction error on the test set.

After several trial runs, we've fixed the kernel class to Gaussian RBFs (it consistently gave the best results).

The training contexts $\mathbf{x}_i$ originally consisted only of the history $\{a_{i-H}, \ldots, a_i\}$ (where H is the age of the oldest sample in the context), but we've found out that adding simple transformations improves the results. The transformation set with the best results turned out to be

$$\mathbf{x}_i = \bigcup_{h=0}^{H} \{a_{i-h},\, a_{i-h}^2,\, \log a_{i-h},\, m_{i,h},\, \log m_{i,h},\, M_{i,h},\, \log M_{i,h}\}$$

where $m_{i,h} = \min\{a_{i-h}, \ldots, a_i\}$ and $M_{i,h} = \max\{a_{i-h}, \ldots, a_i\}$ are moving minima and maxima.

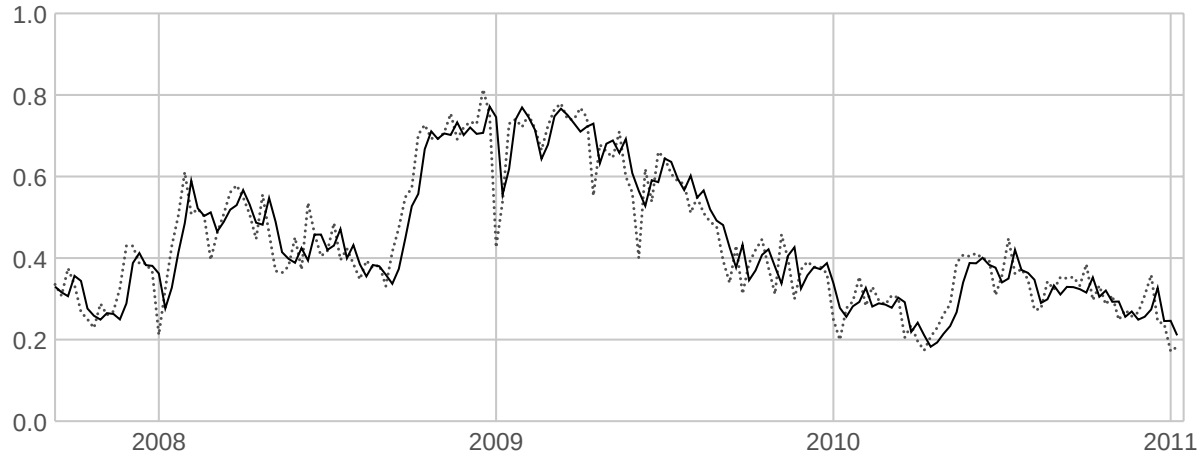


Figure 1 CISS predictions one week into the future, average absolute error = 0.0487. The dotted line denotes the actual value of the index, the solid line denotes the predicted value.

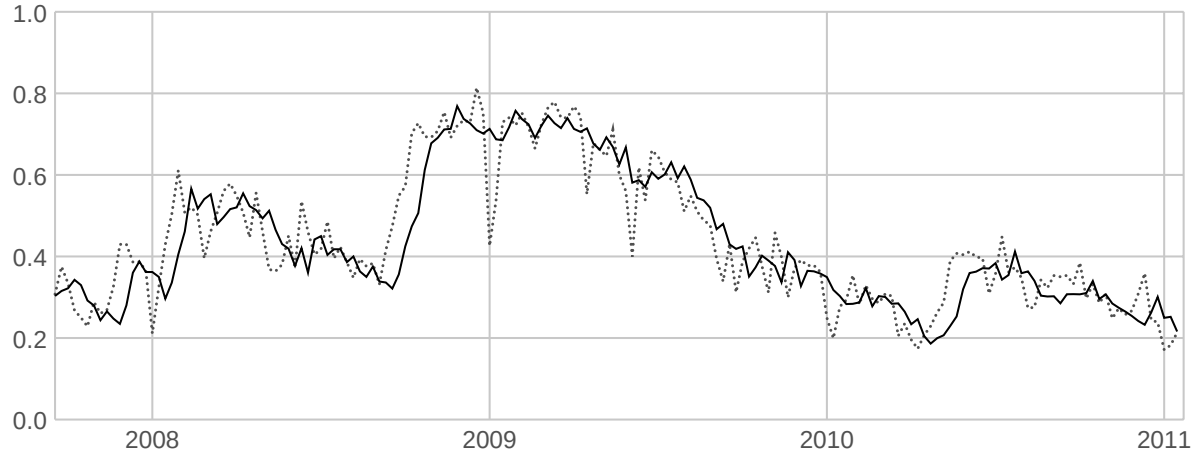


Figure 2 CISS predictions two weeks into the future, average absolute error = 0.0585.

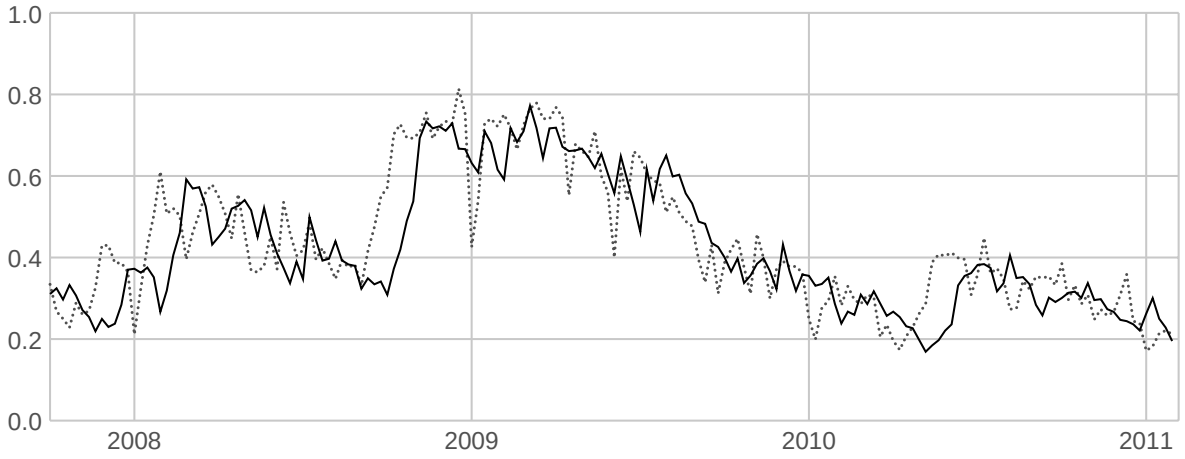


Figure 3 CISS predictions four weeks into the future, average absolute error = 0.0748.

Figure 4 CISS predictions six weeks into the future, average absolute error = 0.0836.

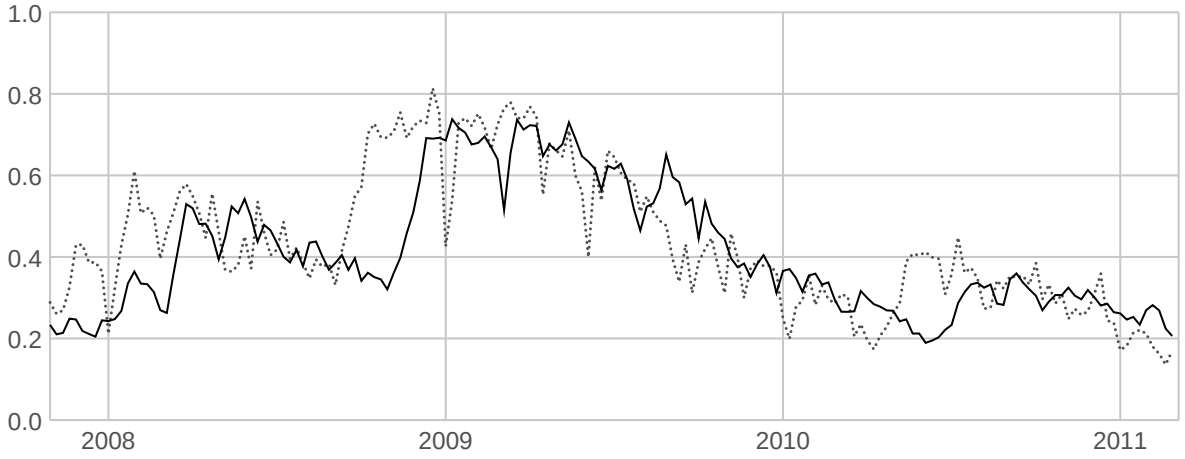


Figure 5 CISS predictions eight weeks into the future, average absolute error = 0.0934.

All components of the context were finally normalized to zero mean and unit variance. Furthermore, it proved beneficial to create contexts only from a subset $\mathcal{S}$ of the full layout. Thus the final parameter set to optimize was $\{\varepsilon, C, \sigma, H, \mathcal{S}\}$.

SVM training and classification was performed with the use of the SVM^{light} library [5].

Figures 1–5 depict the predictions of the best models for each time skip δ. It can be seen that the CISS index can be partially predicted just from its past values. Large behaviour shifts are, however, surprising for the learned model. Incorporating external variables into the model might improve the predictions: this will be a subject of further study.

The most surprising result is that the history length cap H fluctuated between only two and three months. This suggests that data beyond three months ago only confuses the SVM algorithm and that the increased model complexity isn't justified.

4 Conclusion

We've shown the predictability of the CISS index from only its past values for up to eight weeks ahead by the means of a non-linear SVM model. We've optimized the SVM learning parameters and the context layout to minimize prediction error.

In future research we'll try to further improve the prediction accuracy by including additional micro- and macroeconomic time series into the context layout. Based on context lengths of the best obtained models, we conjecture that adding samples beyond three months in the past won't bring any improvements.

Acknowledgements

The author would like to express her thanks to her supervisor Prof. Ing. Miloslav Vošvrda, CSc. and to her colleagues in the ÚTIA AV ČR postgraduate workshop.

References

[1] Balakrishnan, R. et al.: The Transmission of Financial Stress from Advanced to Emerging Economies, *IMF working papers* **09/133** (2009), 1–52.

[2] Burges, C. J. C.: A tutorial on support vector machines for pattern recognition, *Data mining and knowledge discovery* **2** (1998), 121–167.

[3] Cardarelli, R. et al.: Financial stress and economic contractions, *J. Financial Stability* **7** (2011), 78–97.

[4] Holló, D. et al.: CISS – A Composite Indicator of Systemic Stress in the Financial System (2011).

[5] Joachims, T.: Making Large-Scale SVM Learning Practical. In: *Advances in Kernel Methods – Support Vector Learning* (Schölkopf, B., Burges, C., and Smola, A. J., ed.), MIT Press, 1999, 41–56.

[6] Kowalczyk, A.: Maximal margin perceptron. In: *Advances in Large Margin Classifiers* (Smola, A. J., Bartlett, P. L., Schölkopf, B., and Schuurmans, D., eds.), MIT Press, Cambridge, MA, 2000, 75–113.

[7] Laeven, L., and Valencia, F.: Resolution of banking crises: the good, the bad, and the ugly, *IMF working paper* **10/146** (2010).

[8] Mukherjee, S. et al.: Nonlinear prediction of chaotic time series using support vector machines, *Proc. of IEEE NNSP'97*, 1997, 511–520.

[9] Smola, A. J., and Schölkopf, B.: A tutorial on support vector regression, *Statistics and Computing* **14** (2004), 199—222.

[10] Vapnik, V. N.: *The Nature of Statistical Learning Theory*. Springer, 1995.

Data envelopment analysis models with network structure

Josef Jablonský[1]

Abstract. Data envelopment analysis (DEA) is a non-parametric technique for evaluation of relative efficiency of decision making units described by multiple inputs and outputs. It is based on solving linear programming problems. Since 1978 when basic DEA model was introduced many its modifications were formulated. Among them are two or multi-stage models with serial or parallel structure often called network DEA models that are widely discussed in professional community in the last years. The paper presents several approaches for analysis of network production systems and formulates DEA network models for two-stage serial and parallel systems. The presented models are used in a case study that is focused on evaluation of production and profit efficiency of banks in the Czech Republic. The results given by two single DEA models for the first and second stages and then by network model are presented and compared.

Keywords: data envelopment analysis, two-stage models, network models, efficiency

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 can be expressed as the weighted sum of outputs divided by the weighted sum of outputs with weights reflecting the importance of single inputs/outputs v_i, $i = 1,2,\ldots,m$ and u_k, $k = 1,2,\ldots,r$:

$$\theta_q = \frac{\sum_{k=1}^{r} u_k y_{kq}}{\sum_{i=1}^{m} v_i x_{iq}}. \tag{1}$$

Standard CCR input oriented DEA model formulated in [3] Charnes et al. consists in maximization of efficiency score (1) of the DMU_q subject to constraints that efficiency scores of all other DMUs are lower or equal than 1. The linearized form of this model is:

Maximize $\theta_q = \sum_{k=1}^{r} u_k y_{kq}$

subject to (2)

$$\sum_{i=1}^{m} v_i x_{iq} = 1,$$
$$\sum_{k=1}^{r} u_k y_{kj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, \quad j = 1,2,\ldots,n,$$
$$u_k, v_i \ge \varepsilon, \quad k = 1,2,\ldots,r, i = 1,2,\ldots,m.$$

If the optimal value of the model (2) $\theta^*_q = 1$ then the DMU_q is CCR efficient and it is lying on the CCR efficient frontier, otherwise the unit is not CCR efficient. The model (2) is often referenced as primal CCR model. Its dual form is common and its mathematical model is as follows:

[1] University of Economics, Prague, Department of Econometrics, W. Churchill Sq. 4, 13067 Praha 3, Czech Repubic, e-mail: jablon@vse.cz

$$\begin{aligned}
&\text{Minimze} && \theta_q \\
&\text{subject to} && \\
&&& \sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = \theta_q x_{iq}, \qquad i = 1,2,...,m, \\
&&& \sum_{j=1}^{n} y_{kj}\lambda_j - s_k^+ = y_{kq}, \qquad k = 1,2,...,r, \\
&&& \lambda_j \geq 0,\ j = 1,2,\ldots,n,
\end{aligned} \tag{3}$$

where λ_j, $j = 1,2,\ldots,n$ are weights of DMUs, s^-_i, $i = 1,2,\ldots,m$, and s^+_k, $k = 1,2,\ldots,r$ are slack (surplus) variables and θ_q is the efficiency score of the DMU_q which expresses necessary reduction of inputs in order this unit becomes efficient.

The models (2) and (3) are CCR models with input orientation, i.e. they look for reduction of inputs in order to reach the efficient frontier. The output oriented modification of the presented models is straightforward. The BCC models under variable returns to scale assumptions originally presented in [2] by Banker et al. extend the formulation (3) by convexity constraint $\sum_j \lambda_j = 1$. The presented basic DEA models measure efficiency of a transformation of m inputs into r outputs in one stage and under an assumption that all data are deterministic but the production process is often much more complex. That is why various modifications of standard DEA models were formulated in the last years. One stream in this research is represented by network DEA models. They suppose that production process contains several stages with serial or parallel structures.

The paper deals with parallel network models, formulates some of the network models and illustrates their application for two-stage efficiency evaluation of Czech banks. The paper is organized as follows. The next section contains basic formulation of DEA network models. Section 3 presents a simple two-stage model for efficiency evaluation of Czech banks and discusses results of numerical experiments. Final part of the paper summarizes presented results and discusses directions for future research.

2 Network DEA models

The models (2) and (3) measure the relative efficiency of one-stage transformation of m inputs into r outputs. The transformation of inputs into final outputs can be taken as a two- or several-stage process. The inputs of the first stage are transformed into its outputs and all or at least some of these outputs are utilized as inputs of the second stage that are using for production of final outputs. Let us denote the input values of the first stage x_{ij}, $i = 1,2,\ldots,m$, $j = 1,2,\ldots,n$ and the output values of the first stage y_{kj}, $k = 1,2,\ldots,r$, $j = 1,2,\ldots,n$. Supposing that all outputs of the first stage are taken as inputs of the second stage and that the final output values are z_{lj}, $l = 1,2,\ldots,p$, $j = 1,2,\ldots,n$. Two-stage DEA models are widely analyzed and discussed within professional community. Theoretical issues can be found e.g. in [7]. Among numerous case studies can be mentioned papers [5] and [8]. One of the first two-stage serial DEA models is formulated in [6]. Below is its dual formulation with constant returns to scale assumption and input orientation:

$$\begin{aligned}
&\text{Minimize} && \theta_q \\
&\text{subject to} && \sum_{j=1}^{n} x_{ij}\lambda_j \leq \theta_q x_{iq}, && i = 1,2,...,m, \\
&&& \sum_{j=1}^{n} y_{kj}\lambda_j - \sum_{j=1}^{n} y_{kj}\mu_j \geq 0, && k = 1,2,...,r, \\
&&& \sum_{j=1}^{n} z_{lj}\mu_j \geq z_{lq}, && l = 1,2,...,p, \\
&&& \lambda_j \geq 0, \mu_j \geq 0, && j = 1,2,\ldots,n,
\end{aligned} \tag{5}$$

where λ_j and μ_j, $j = 1,2,\ldots,n$, are weights of the DMUs in the first and second stage, and θ_q is efficiency score of the DMU_q. The efficiency measure of the model (5) is always lower or equal to 1 and it is possible to simply prove that it is a product of efficiency measures of two single stages given by model (3) with constant returns to scales - see e.g. [6] or [1]. Target values for inputs, intermediate characteristics and final outputs of the ineffi-

cient DMUs, i.e. characteristics of virtual units, can be given as linear (convex) combination of DMUs using their optimal weights λ^*_j and μ^*_j, $j = 1,2,\ldots,n$.

Another formulation of two-stage DEA model under constant returns to scale assumption is given in [4]. Their formulation follows:

$$
\begin{aligned}
\text{Minimize} \quad & \theta_q - \phi_q \\
\text{subject to} \quad & \sum_{j=1}^{n} x_{ij}\lambda_j \le \theta_q x_{iq}, && i = 1,2,\ldots,m, \\
& \sum_{j=1}^{n} y_{kj}\lambda_j \ge \tilde{y}_{kq}, && k = 1,2,\ldots,r, \\
& \sum_{j=1}^{n} y_{kj}\mu_j \le \tilde{y}_{kq}, && k = 1,2,\ldots,r, \\
& \sum_{j=1}^{n} z_{lj}\mu_j \ge \phi_q z_{lq}, && l = 1,2,\ldots,p, \\
& \theta_q \le 1, \phi_q \ge 1, \\
& \lambda_j \ge 0, \mu_j \ge 0, && j = 1,2,\ldots,n,
\end{aligned}
\tag{6}
$$

where λ_j and μ_j, $j = 1,2,\ldots,n$, are weights of the DMUs in the first and second stage, θ_q and φ_q efficiency scores of the DMU_q in the first and second stage and $\tilde{y}_{kq}$ are variables to be determined. The DMU_q is recognized as efficient by model (4) if the efficiency scores in both stages are $\theta_q = 1$ and $\varphi_q = 1$ respectively, and the optimal objective value of the presented model is 0. Target values for inputs, intermediate characteristics and final outputs of the inefficient DMUs can be in the same way as in the previous model. The inefficient units in model (6) can be ranked relatively by the following geometric average efficiency measure:

$$e_q = (\theta_q / \varphi_q)^{1/2}. \tag{7}$$

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 CZ	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 – efficiency evaluation of Czech banks

The models (5) and (6) are models that measure efficiency of two-stage serial production processes under the assumption of constant returns to scale. They can be modified for variable returns to scale by adding convexity constraints for both sets of weights λ_j and μ_j, $j = 1, 2,\ldots, n$. The models (5) and (6) can be simply generalized for processes with more than two stages – see e.g. [1].

3 Efficiency evaluation of Czech banks

Applications of two- or multiple-stage DEA models are numerous. 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. They are taken from public financial statements of the banks for year 2010. The first stage evaluates production efficiency and the second stage is profit efficiency. The following inputs, intermediate characteristics (outputs of the first stage and inputs of the second one) and final outputs are taken into account (source values of the characteristics for all 20 banks are presented in Table 1):

Inputs:
- Equity of the bank in millions of CZK (Czech crowns),
- Number of full time employees (FTE).

Intermediate characteristics (outputs of the first stage and inputs of the second one):
- Deposits in millions of CZK,
- Credits in millions of CZK.

Final output:
- Profit in millions of CZK.

Bank	**1st stage VRS-I**	**2nd stage VRS-O**	**Overall Geomean**	**Model (5) VRS-I**	**Model (5) VRS-O**	**Model (6) VRS - e_q**
(1)	(2)	(3)	(4)	(5)	(6)	(7)
CS	1.000	0.993	0.997 (5)	0.0836 (19)	0.9930 (4)	0.2890 (15)
Citibank	1.000	1.000	1.000 (1)	1.0000 (1)	1.0000 (1)	1.0000 (1)
ČSOB	1.000	0.946	0.972 (6)	0.1709 (14)	0.9446 (5)	0.4190 (7)
GE Money	0.464	0.407	0.434 (14)	0.0485 (20)	0.2691 (7)	0.4834 (5)
Hypotecní Bank	1.000	1.000	1.000 (1)	0.1716 (13)	0.3286 (6)	0.4179 (8)
ING CZ	1.000	0.761	0.873 (7)	0.3832 (7)	0.1773 (9)	0.4210 (6)
KB	1.000	1.000	1.000 (1)	1.0000 (1)	1.0000 (1)	1.0000 (1)
LBBW Bank	0.150	0.046	0.084 (20)	0.0943 (18)	0.0045 (20)	0.0667 (20)
Raiffeisenbank	0.968	0.207	0.447 (13)	0.1336 (16)	0.1688 (10)	0.4109 (10)
UniCredit	1.000	0.243	0.493 (10)	0.3127 (9)	0.2427 (8)	0.4849 (4)
Volksbank CZ	0.159	0.132	0.145 (19)	0.0975 (17)	0.0468 (17)	0.2163 (17)
Wustenrot	0.399	0.127	0.225 (17)	0.1835 (12)	0.0429 (18)	0.2071 (18)
Commerzbank	0.716	0.182	0.361 (15)	0.4314 (6)	0.0656 (16)	0.2561 (16)
Czech Export Bank	1.000	0.043	0.207 (18)	0.4503 (5)	0.0428 (19)	0.1684 (19)
Czech-Moravian	0.302	0.823	0.499 (9)	0.2355 (10)	0.1584 (12)	0.3980 (11)
PPF Bank	0.688	0.793	0.739 (8)	0.4581 (4)	0.1525 (13)	0.3905 (12)
Blue pyramid	0.373	0.213	0.282 (16)	0.1392 (15)	0.1133 (15)	0.3367 (14)
Raiffeisen SS	1.000	0.229	0.478 (11)	0.3620 (8)	0.1502 (14)	0.3503 (13)
CS SS	0.639	0.347	0.471 (12)	0.2227 (11)	0.1716 (11)	0.4143 (9)
Wustenrot hyp	1.000	1.000	1.000 (1)	1.0000 (1)	1.0000 (1)	1.0000 (1)

Table 2 Efficiency scores given by DEA models

The results and comparison of presented models for all DMUs are included in Table 2. All the results are given by models under the assumption of variable returns to scale which seems to be more suitable for evaluation purposes. The table contains the following information:

- Column (1) – name of the bank.
- Column (2) – efficiency score of the first stage given by standard BCC input oriented model. As presented, half of the banks are determined as efficient by this model in the first stage.
- Column (3) – efficiency score of the second stage given by BCC output oriented model. As the efficiency scores given by output oriented models are greater or equal 1, their reciprocal values are presented because it is more understandable for decision makers. Only 4 banks are recognized as efficient by this model.
- Column (4) – geometric average of efficiency scores in the previous two columns and ranking of DMUs by this value (in parenthesis). Four units (Citibank, Hypotecni Bank, KB, and Wustenrot hyp) are efficient according to this measure.
- Column (5) – efficiency scores computed by two-stage model (5) with input orientation and variable returns to scale assumption. The results of this model are quite surprising. Some of the DMUs efficient or nearly efficient by two single stages models are rated very badly in this model. It holds especially for first DMU (CS). The reason is its low final output (profit) comparing to its inputs (equity and FTE) and inputs and profit of other units of the decision set.
- Column (6) – the same characteristics as in the previous column but the model (5) with output orientation is used. The efficiency scores are given as their reciprocal values as in column (3). The first DMU is almost efficient by this model and it is immediately after the first three efficient units with maximum level of efficiency. Differences in rankings of DMUs by input- and output oriented models are significant and they are hardly possible to explain.
- Column (7) – efficiency scores given by model (6) with variable returns to scale transformed using formula (7). The model (6) evaluates the efficiency in two stages simultaneously and input orientation in the first stage and output orientation in the second one is applied.

The results (efficiency score) of the model (5) are questionable. The efficiency score is a product of two efficiency scores from both stages with the same orientation and the same returns to scale assumption. This "overall" efficiency score can be simply used for ranking of DMUs but has no direct explanation. A special attention must be given to the orientation of the model because when applying the model the results are quite different depending on the orientation. The model (6) evaluates the two-stage efficiency simultaneously without necessity to choose a model orientation. It is its advantage but the interpretation of the results is problematic as in the previous case. A promising research direction consists in formulation of two- or multi-stage serial DEA models based on slack based DEA models. The model formulated in [9] by Tone seems to be an ideal starting point for experiments in this area.

4 Conclusions

Evaluation of efficiency of network production systems is a very complex task that can be solved using network DEA models. This class of models is widely discussed within professional community in the last years and many papers are published with a main focus on theoretical aspects and/or applications in this field. This paper is focused on a simplest system which is two-stage serial model. Two of main two-stage serial DEA models are formulated and applied on a real data set regarding banks operated in the Czech Republic. Efficiency of banks is evaluated by two-stage model where the first stage takes into account production efficiency and the second one profit efficiency. The conclusions given by applied models show quite significant differences in ranking of banks. It is difficult to explain the results (efficiency scores) and it is one of the most important disadvantages of current two-stage models. Further research will be focused on formulation of special network models with better stability of results and their clearer interpretation. Other interesting area is analysis of super-efficiency of DMUs in network models.

Acknowledgements

The research is supported by the Czech Science Foundation, project no. 402/12/1387, and by IGA FIS VSE Praha, project no. F4/18/2011.

References

[1] Ashrafi, A., Jaafar, A.B., Lee, L.S. and Abu Bakar, M.R., The efficiency measurement of parallel production systems: a non-radial DEA model. *Journal of Computer Science*, **7**, 5, 749-756 (2011).

[2] Banker, R. D., Charnes, A. and Cooper, W. W., Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science*, **30**, 9, 1078–1092 (1984).

[3] Charnes, A., Cooper, W. W. and Rhodes, E., Measuring the efficiency of decision making units. *European Journal of Operational Research*, **2**, 6, 429–444 (1978).

[4] Chen, Y., Liang, L. and Zhu, J., Equivalence in two-stage DEA approaches. *European Journal of Operational Research*, **193**, 600-604 (2009).
[5] Jablonsky, J. A two-stage AHP/DEA model for evaluation of efficiency. In: T. Subrt (Ed.), *Mathematical Methods in Economics 2009*, University of Life Sciences Press, Praha, 159 - 164 (2009).
[6] Kao, C. and Hwang, S.N., Efficiency decomposition in two-stage data envelopment analysis: An application to non-life insurance companies in Taiwan. *European Journal of Operational Research*, **185**, 1, 418-429 (2008).
[7] Liang, L., Li, Z. Q., Cook, W.D. and Zhu, J., Data envelopment analysis efficiency in two-stage networks with feedback. *IIE Transactions*, **43**, 5, 309-322 (2011).
[8] Paradi, J.C., Rouatt, C. and Zhu, H., Two-stage evaluation of bank branch efficiency using data envelopment analysis. *Omega*, **39**, 1, 99-109 (2011).
[9] Tone, K.: A slack-based measure of super-efficiency in data envelopment analysis. *European Journal of Operational Research*, **143**, 32-41 (2002).
[10] Zhu, J. *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets and DEA Excel Solver*. Boston: Kluwer Academic Publishers (2003).

Territory decomposition parameters of distribution tasks

Marta Janáčková[1], Alžbeta Szendreyová[2]

Abstract. Number of candidates for the location of centers and their locations affects the quality of the location problems solution. When the territory is evenly partitioned into the squares, the edge size impacts both the population density size in individual squares and numbers of qualitatively different squares on the border of the solved area. This paper deals with the relationship between the parameters.

Keywords: *p*-median, uncapacitated location problem, service center, private service systems, public service systems.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Both public and private service systems are involved in finding out a solution of many tasks. The private service systems take into account the maximization of the profit. On the other hand, the public service systems have to allow the approach to the common service for each customer (health services, education) in spite of disadvantageous conditions. The satisfaction of customer's demands is limited by many factors, such as financial resources, influence on the geographic environment, etc. The quality of the service can be evaluated by several criteria. The service systems properties are described in [3].

Uncapacitated location problem with limited number of service places is one of the service system tasks (*p*-median). The quality criterion of service provision can be the minimization of the maximal distance between a customer and its closest center. Another criterion is the minimization of the travel kilometers in accordance with the optimal average accessibility of service. The size of the set of candidates for the centers location impacts on the quality of the task result. The location of the candidates, the distribution of possible center locations [4], the transport infrastructure and other parameters impact the result as well. Some parameters mutually influence each other. Um Jaegon [2] deduced the relationship whereby the suitable number of candidates is in proportion to the population density in the region.

This paper deals with the possibilities how to achieve the similar status, if we solve the *p*-median problem by using our software. That means how to select the sets and candidates for possible centers that they evenly cover the serviced area and fulfill the specified proportional relationship.

2 Selection of candidates

Um Jaegon describes in [2] that the number of candidates for the service center location meets the expression $D=const.\rho^{2/3}$, where D is the density of candidates and ρ is the population density in the checked component. Our task is to place m candidates for the location. They should to fulfill the mentioned formula as good as possible.

For obtained set of candidates the *p*-median problem is solved, or the impact on the quality of service accessibility is evaluated. In this paper we consider that the quality of service is the total number of traveled kilometers.

We can select the set of candidates randomly, or the selection can be based on different criteria. In previous studies we found that for *p*-median we got better results, if the set contained more candidates. The results were better, if the candidates were placed in large cities as well. On the other hand, the necessity to ensure an equitable access of customers to the service makes us to deploy the candidates evenly throughout the served territory. Therefore, we divide the territory into smaller regular units and we choose the candidates for location from each group. The division of the territory to the individual components can be done in different ways. In some cases it is appropriate to maintain the compactness of the road network, or the borders of historic areas, respectively the division by regions. In the current proposal of the service system we use automated division of the area into squares.

The program for regular division of the area was created as a result of task solved in the thesis [1]. We used in our study its revised application. It allows to cover the serviced area with grid and to include the nodes of the

[1] University of Žilina, Univerzitná 1, 010 26 Žilina, Slovakia, Marta.Janackova@fstroj.uniza.sk.
[2] University of Žilina, Univerzitná 1, 010 26 Žilina, Slovakia, Alzbeta.Szendreyova@fri.uniza.sk.

network into each square of the grid. After the selection of the criteria, the program selects some network nodes and includes them into the set of candidates. The properties of the grid are determined by entering the side length of the square. It is possible to determine the square population and to calculate the density of the population.

It is clear that the side length of the square is in reverse proportion to the number of the squares. Since each square contains the same number of candidates, the number of squares is in direct proportion to the number of candidates. The side length of the square influences the assignment of nodes and hence the density of the population in each square. The question is how to choose the size of components, i.e. how to determine the side length of the square so that the population density in each square will not be very different. The side length of the square changes the boundaries of the components and also the assignment of the nodes to the components. The population density in the components (squares) depends on the side length of the square and on the geographical location of nodes as well. It is hard to pre-calculate how the density in the components will change. We handle the automated division of the area to the squares. It makes easy to change the number of the components and to process a large number of tasks. We use the obtained data as a basis for determining the number of candidates in each square according to the expression mentioned above.

The population densities in the squares of the grid are different due to geographical shape of the Slovakia territory. We expect that the density will align with increasing size of the square. Because each grid contains from tens to hundreds squares, we separate the squares into classes according to the population densities. We distribute the estimated number of candidates to the squares in proportion to the mentioned formula. A mean value from the interval of each class is used as the population density. The dependency on relationship between the number of candidates and the population density is not linear. The calculated values are not integer and we use them as a basis for determining the set of candidates. This set can be interactively adjusted. We compare the results and find out for which side length of the square will be the location of candidates according to the densities similar to their uniform distribution.

The results of the sorting and dividing a set of candidates into the squares are given in the chapter with the experiments.

3 *p*-median problem

To verify the properties of a decomposition of a candidates set, we solve the *p*-median problem. This is a known location problem with a limited number of possible center locations. Let I is the set of possible candidates for the center locations. Customers are placed at the network nodes J. Every municipality $j \in J$ has b_j customers. The segment between i and j is evaluated as a distance d_{ij} for each possible location $i \in I$ and each dwelling place $j \in J$. The *p*-median problem solves the location of p centers at some nodes from the set I. These p centers have to serve each customers $j \in J$ in order to minimize the number of traveled kilometers. We assume that all $b_j, j \in J$ customers will be served from exactly one center $i \in I$. In this case, the coefficient d_{ij} can be modified by product of distance d_{ij} and the number of customers b_j of the place $j \in J$. So, $c_{ij} = b_j.d_{ij}$ for $i \in I$ and $j \in J$.

The decision on locating or not locating a center i at a place $i \in I$ will be modeled by a variable y_i which takes the value of 1 if the center is located at place i and it takes the value of 0 otherwise. The decision on allocation of the customer from node j to the center at the place i is modeled by a variable z_{ij}. It takes the value of 1 if the customer j will be served from the center i and takes the value of 0 otherwise. A model of the problem follows:

$$\textit{Minimize} \quad \sum_{i \in I} \sum_{j \in J} c_{ij} z_{ij} \tag{1}$$

$$\textit{Subject to} \quad \sum_{i \in I} z_{ij} = 1 \qquad \text{for } j \in J \tag{2}$$

$$z_{ij} \le y_i \qquad \text{for } i \in I,\ j \in J \tag{3}$$

$$\sum_{i \in I} y_i \le p \tag{4}$$

$$y_i \in \{0,1\} \qquad \text{for } i \in I \tag{5}$$

$$z_{ij} \in \{0,1\} \qquad \text{for } i \in I,\ j \in J \tag{6}$$

The used coefficients have the following meanings:

c_{ij} … evaluation of the assignment;
p … required number of centers;
I … the set of possible service center locations;
J … the set of customers.

Expression (1) is the objective function. Constraints (2) guarantee that a customer from dwelling place $j \in J$ is assigned to the exactly one center $i \in I$. The constraints (3) ensure that the variable y_i takes the value of 1, when at least one customer is assigned to the center at place i. The constraint (4) put the limit p on the number of location.

4 Numerical experiments

The whole problem was solved on the road network of the Slovak Republic. The network consists of 2916 dwelling places with potential customers. The position of places and the numbers of inhabitants (potential customers) are known. The area of the Slovak Republic is about 49000 km^2.

The area was divided into squares on the basis of the square length. The length of squares acquires the values from 5 to 30 kilometers. A lot of squares did not include any dwelling place. In these squares no candidates were located, so it did not affect the result. Table 1 lists the number of squares according to the side length of the square in the grid.

a [km]	V	O	O [%]	S [km^2]
5	3440	1492	43.37	37300
10	860	516	60.00	51600
15	406	252	62.07	56700
20	220	152	69.09	60800
25	144	101	70.14	63125
30	105	73	69.52	65700

a-side length of the square, V-number of all squares, O-number of the inhabited squares, S-area of the inhabited squares

Table 1 Decomposition of the area

For better comparability we also present the percentage and the area of the inhabited squares. This area is not the same as the area of Slovak Republic because some of the squares are situated in the foreign territory and some squares are in inactive (without population) squares.

Squares, which were situated on the border of the area, had only a small number of dwelling places and there were also very small population in these squares. The densities in these squares were much smaller as the density of the squares with big cities. For each side length of the square we found the minimum and maximum densities of all active squares. We chose a smaller width of the class at the beginning of the interval for to separate the active squares with negligible density. Six classes for each grid were determined. In the classes were squares with the population density up to 1 percent of maximum density and further up to 5%, 25%, 50% and finally up to 100% of the maximum density of all the grid squares.

For each side length of the square, i.e. for each grid, we found the frequency of the classes, and we express the relative frequency with respect to the number of active squares of the grid. These data are shown in Table 2.

Density	**< 1%**	**1-5%**	**5-25%**	**25-50%**	**50-75%**	**<100%**
a=5; min=1.12; max=5528; number of squares=1493						
i	55.29	276.44	1382.22	2764.44	4146.66	5528.88
b_i	707	664	99	14	6	3
g_i	47.35%	44.47%	6.63%	0.94%	0.40%	0.20%
a=10; min=0.72; max=2740; number of squares=516						
i	27.40	137.01	685.07	1370.15	2055.22	2740.29
b_i	130	300	76	8	0	2
g_i	25.19%	58.14%	14.73%	1.55%	0.00%	0.39%

a=15; min=0.12; max=1423; number of squares=252						
i	14.23	71.17	355.87	711.74	1067.60	1423.47
b_i	31	123	89	7	1	1
g_i	12.30%	48.81%	35.32%	2.78%	0.40%	0.40%
a=20; min=0.53; max=695; number of squares=152						
i	6.95	34.75	173.77	347.55	521.32	695.09
b_i	10	36	89	13	2	2
g_i	6.58%	23.68%	58.55%	8.55%	1.32%	1.32%
a=25; min=1.13; max=700; number of squares=101						
i	7.00	34.98	174.90	349.79	524.69	699.58
b_i	10	16	67	6	1	1
g_i	9.90%	15.84%	66.34%	5.94%	0.99%	0.99%
a=30; min=0.20; max=445; number of squares=73						
i	4.45	22.25	111.23	222.46	333.69	444.92
b_i	3	6	46	15	2	1
g_i	4.11%	8.22%	63.01%	20.55%	2.74%	1.37%

i-upper limit of the interval i according to population density, b_i - number of squares in a class, g_i- percentage of b_i

Table 2 Classes according to the density

For each of decomposition, we created sets of candidates. They contained from 1 to 6 candidates in each inhabited square. The candidates were placed in communities with the largest population. We used the frequency of the classes and mean values of the densities to calculate the values (proportionality criteria) for each grid using the expression:

$$m_i = \frac{m.\rho_i^{2/3}}{\sum_{i=1}^{6} b_i.\rho_i^{2/3}} \tag{7}$$

where b_i is the frequency of the i-th class and ρ_i is the middle of the interval of i-th class. The closest integer value to m_i is a proposal to the number of the candidates for each square of i-th class. These values are listed in Table 3. The last value in Table 3 shows the percentage of candidates placed in the squares considering to the total number of candidates in the grid and according to this criterion.

Density	**< 1%**	**1-5%**	**5-25%**	**25-50%**	**50-75%**	**<100%**
a=5; min=1.12; max=5528; number of squares=1493						
m_i	0.34	1.13	3.32	6.11	8.58	10.74
h_i	16.26%	50.42%	21.98%	5.73%	3.45%	2.16%
a=10; min=0.72; max=2740; number of squares=516						
m_i	0.25	0.83	2.42	4.46	6.26	7.84
h_i	6.31%	48.11%	35.63%	6.91%	0.00%	3.04%
a=15; min=0.12; max=1423; number of squares=252						
m_i	0.17	0.56	1.65	3.03	4.26	5.34
h_i	2.10%	27.49%	58.17%	8.43%	1.69%	2.12%
a=20; min=0.53; max=695; number of squares=152						
m_i	0.11	0.38	1.10	2.03	2.85	3.57
h_i	0.75%	8.92%	64.51%	17.36%	3.75%	4.70%

a=25; min=1.13; max=700; number of squares=101						
m_i	0.12	0.38	1.12	2.06	2.89	3.62
h_i	1.15%	6.05%	74.12%	12.23%	2.86%	3.59%
a=30; min=0.20; max=445; number of squares=73						
m_i	0.09	0.30	0.86	1.59	2.24	2.80
h_i	0.37%	2.43%	54.49%	32.73%	6.14%	3.84%

m_i-corresponding number of candidates per square of a class, h_i - percentage of candidates pertaining to the class

Table 3 Proposed number of candidates according to population density

We solved the *p*-median task with *p*=100 for each decomposition, which was determined by the side length of the square and by sets of candidates. The objective function values depending on the side lengths of the square are presented in Table 4.

a / *K*	***a*=5**	***a*=10**	***a*=15**	***a*=20**	***a*=25**	***a*=30**
1	31954028	33414347	34346567	36105439	45110994	---
2	38181714	33106788	32192284	33007281	34476130	36640763
3	48562293	33391232	31985927	32716026	33327815	32931305
4	---	34151736	31934509	31948971	33057442	32654166
5	---	36513845	32351365	31948971	32247703	32003904
6	---	40797022	33427365	31933226	32200492	31961588

a-side length of the square, *K*-number of candidates per square

Table 4 Values of the objective function

To solve the *p*-median task we used software product, which was implemented in the Department of Transport Networks, University of Žilina. It is based on the branch and bound method. Condition (3) of the model was ensured by Lagrange relaxation. Experiments were performed on a laptop with Windows Vista, 3.00 GB of RAM and 2.4 GHz processor.

5 Conclusions

The results show that with decreasing side length of the square, both the number of active squares and the maximum population density vary considerably. When the side length of the square is small (*a*=5), only a negligible amount of squares has population density greater than 5% of maximum. On the contrary, for *a*=30, only 12% of squares has a density less than 5% of maximum. In spite of these differences, the relative frequency with the square density of up to 25% of the maximum density reaches for all types of grid comparable values. While for *a*=5, *a*=10 it is 98%, for *a*=15 it is 96%, for *a*=20 reaches 89%, for *a*=25 to 92% and finally for *a*=30 reaches 75%.

Assignment the number of candidates according to the population density of the square is as follows: for *a*=5 proposal contains 0 to 10 candidates per square. When the side length a increasing the number of candidates decreases. For *a*=30 it is from 0 to a maximum of 3 candidates per square. If we sum up the squares in which the criterion recommends to place approximately 1-2 candidates (m_i is close to this value), then the amount for each of the decomposition exceeds 80% of the number of candidates.

Among the best results from the *p*-median problem for each of decompositions, there are no significant differences. If necessary, the result of automatic selection of a set of candidates according to the specified parameters can be adjusted manually.

Acknowledgements

This work 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".

We would like to thank to "Centre of excellence in computer sciences and knowledge systems (ITMS 26220120007) for built up infrastructure, which was used.

References

[1] Cudrák, P.: *Diploma thesis*, ŽU 2011

[2] Um, J. at al: Scaling laws between population and facility densities. In: *Proceedings of the National Academy of Sciences of the USA*, Vol. 106, No. 34, p. 14236-14240, August 2009

[3] Janáček, J. a kol.: *Navrhovanie územne rozľahlých obslužných systémov (Designing of spatially extensive service systems)*, monograph, p. 404, EDIS - publishers ŽU, Žilina, 2010

[4] Janáčková, M., and Szendreyová, A. : First Aid Stations Location Problem and an Impact of the Set of Candidates on the Recourse of a Patient. In: *Proceedings of the SOR ´09, Nova Gorica, Slovenia*, 2009, p. 357-360, ISBN 978-961-6165-30-3

DEA as a tool for bankruptcy assessment: the agribusiness case study

Jitka Janová[1], Jan Vavřina[2], David Hampel[3]

Abstract. Bankruptcy assessment provides valuable information for the governments and investors to base their decisions in order to prevent possible financial losses. Data envelopment analysis (DEA) has generally been used to assess relative efficiency of decision making units. Recently, several approaches have appeared that reformulate DEA as a bankruptcy prediction tool. However, only several studies have been published on evaluation of suggested approaches and, if so, the focus was given to manufacture industry and IT industry only.

In this paper we discuss the possibility of application of recent results in DEA bankruptcy prediction models in a specific field of agribusiness. Using the AMADEUS database, the Czech Farm Accountancy Data Network (FADN CZ) and the financial statements issued in the Czech Business Register, we collect primary data set on the Czech agriculture firms financial performance, apply the DEA based model, evaluate the results obtained and discuss the prediction power of the approaches in the agriculture industry.

Keywords: DEA, agriculture, bankruptcy assessment

JEL classification: C61
AMS classification: 90C90

1 Introduction

Bankruptcy assessment has already been an intensively studied problem, however, after recent financial crisis, the urgent need for unchallenged techniques evaluating properly the financial health of enterprises has even arisen. Apart form testing the applicability of classical Altman's bankruptcy prediction model, further approaches have been suggested that apply or modify the well established methods for bankruptcy assessment usage. Let us mention the logistic regression and the discriminant analysis that dominate the literature (for review see [9]) or more recent methods employing e.g a chaos approach [10] or neural networks applications [11]. However, the reliability of techniques can still be increased, hence the research in the field focuses on improving the known methods or suggesting and validating the new ones. One of the most distinguished methods from those recently arisen for the bankruptcy assessment is the DEA (data envelopment analysis) approach. Premachandra [12] introduces in 2009 (DEA) as a non-parametric approach for analysing enterprises performance and suggests that this approach can be used as a help for bankruptcy assessment. In our contribution we focus on introductory evaluation of this recently suggested technique in the specific field of agribusiness.

Originally, DEA was developed to analyze efficiencies of decision making units (see e.g. [6]) and the method has been applied in many different business branches. As for the agribusiness applications, DEA was applied only scarcely until the end of 20th century (see [5]). However, since then, DEA became a favorite tool for farm efficiency measurement and number of local studies applying the DEA methodology in agriculture have been elaborated (see e.g. [3]). DEA was used not only to evaluate the efficiencies of agriculture enterprises, but also the determinants of efficiency in agribusiness were studied ([2]) and some studies discussing DEA with respect to the problem of sampling variations existence were provided

[1]Mendel University in Brno/Department of statistics and operational analysis, Zemědělská 1, Brno 613 00, e-mail: janova@mendelu.cz
[2]Mendel University in Brno/Department of business economics, Zemědělská 1, Brno 613 00, e-mail: vavrina@node.mendelu.cz
[3]Mendel University in Brno/Department of statistics and operational analysis, Zemědělská 1, Brno 613 00, e-mail: qqhampel@node.mendelu.cz

for the specific field of agriculture (e.g. [1]). Especially, the local agribusiness DEA studies were carried out for which the national FADN databases were employed as data sources and this implied also the selection of DEA input and output variables used in particular studies (for Czech agribusiness problem see [7]). While the indices have already been investigated for the use in agribusiness before, application of recently theoretically developed DEA bankruptcy prediction represents a novel approach in the field. This new approach for financial distress prediction has not been studied for the agriculture industry so far.

2 Methods

Premachandra [12] selects for bankruptcy assessment the additive DEA model [4] which evaluates the relative efficiency of the specific oth firm as follows:

$$\begin{aligned} \max \quad & es^- + es^+ \\ \text{subject to :} \quad & X\lambda + s^- = x_o, \\ & Y\lambda - s^+ = y_o, \\ & e\lambda = 1, \\ & \lambda \geq 0, \\ & s^- \geq 0, \\ & s^+ \geq 0. \end{aligned} \tag{1}$$

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 the weight vector.

Note that the additive model allows negative values in inputs and outputs, which is useful in bankruptcy assessment where financial ratios (often negative) enter the calculations and, moreover, the additive model incorporates both input and output slacks in the efficiency measurement and the efficiency of a specific decision making unit is determined by examining slacks only. This feature seems comfortable for users who need not examine both DEA efficiency score and slacks.

For bankruptcy assessment, the role of inputs and outputs in (1) is played by financial variables (concretely by financial ratios), but as usually in DEA, the concrete selection is determined by a DEA user. In our study we will follow the input output choice suggested in [12], but note that using ratios in DEA analysis is currently being a matter of scientific discussion (see e.g. [8]). Respecting the data availability the final set of in/outputs has been slightly modified using *book value of total debt* instead of *interest expense*:

1. inputs
   - CFTA = cash flow/total assets
   - WCTA = working capital/ total assets
   - ETA = EBIT/ total assets
   - ER = EBIT/revenues
   - ETD = book value of equity/ book value of total debt
   - NITA = net income/ total assets
   - CATA current assets/ total assets
2. outputs
   - TDTA = total debts/ total assets
   - CLTA = current liabilities/ total assets.

Unlike in the conventional DEA based production analysis, where productive performers consist an efficiency frontier and insufficient performers exist within the production possibility set, in the approach

considered here the frontier is "bankruptcy frontier". This means that the frontier contains the poor performers - the bankrupt firms, while the healthy firms are expected to exist inside a "bankruptcy possibility set". Hence, solving the additive model (1) for each firm we classify the firm based on wheter all the slacks are zero on optimality of (1). If all slacks are zero, the firm is on the bankruptcy frontier. Otherwise (at least one slack positive), the firm is not on the bankruptcy frontier.

3 Results

The data on 54 bankrupt and 21 healthy Czech agricultural firms were taken from 2007-2010 FADN and AMADEUS databases. The two sided Wilcoxon rank sum test for the median values of chosen inputs and outputs was applied in MATLAB to indicate significant difference between the bankrupt and non-bankrupt firms. The results can be seen in Table 1 where the medians for bankrupt ($median\, b$) and non-bankrupt ($median\, non-b$) firms are summarized. In line h there is the result of testing the null hypothesis that data in the vectors composed from bankrupt and non-bankrupt firm's financial characteristics are independent samples from identical continuous distributions with equal medians against the alternative that they do not have equal medians (h = 1 indicates a rejection of the null hypothesis at the 5% significance level, h = 0 indicates a failure to reject the null hypothesis at the 5% significance level). Since for all the in/outputs except CATA the medians were significantly different for bankrupt and non-bankrupt firms, we decide to choose the following final set of inputs and outputs that are appropriate in clasifying the agriculture firms as bankrupt and non-bankrupt:

1. inputs: CFTA, WCTA, ETA, ER, ETD, NITA
2. outputs: TDTA, CLTA.

in/output	TDTA	CLTA	CFTA	WCTA	ETA	ER	ETD	NITA	CATA
median b	1.114	0.511	-0.153	0.076	-0.073	-0.098	-0.104	-0.072	0.481
median non-b	0.275	0.115	0.202	0.223	0.194	0.158	2.642	0.166	0.517
h	1	1	1	1	1	1	1	1	0

Table 1: Financial characteristics for bankrupt and non-bankrupt firms

The capability of DEA approach in evaluating bankruptcy of agriculture enterprises is tested by using 54 different samples, each containing one bankrupt and several healthy firms. Sample sizes were 11, 16 and 21 and we had 20 samples containing 10 healthy firms, 17 samples containing 15 healthy firms and 17 samples containing 20 healthy firms. The healthy firms were randomly selected from the initial data set while each of the bankrupt firms was contained just in one of the 54 samples. This experiment enables to assess the strength of DEA in identifying a single bankrupt firm from other firms in a sample. The calculations were run in MATLAB using a script for cycle solution of DEA that employs the code [13].

	appeared in frontier F	not appeared in frontier NF	total
No. of bankrupt firms B	52	2	54
No. of non-bankrupt firms NB	214	581	795
total	266	583	849

Table 2: Summary of the DEA results

As Table 2 shows, 52 of 54 bankrupt agriculture firms appeared in the bankruptcy frontier and in 214 of total 795 cases the healthy firm appeared in bankruptcy frontier as well. Let us compute the four probabilities:

P_1 = the number of bankrupt firms on the bankruptcy frontier divided by the total number of bankrupt firms

P_2 = the number of bankrupt firms not on the bankruptcy frontier divided by the total number of bankrupt firms

P_3 = the number of non-bankrupt firms not on the bankruptcy frontier divided by the total number of non-bankrupt firms -

P_4 = the number of non-bankrupt firms on the bankruptcy frontier divided by the total number of non- bankrupt firms.

Using the values from our experiment (see Table 2) we obtain the rates

$$P_1 = 0.96,\ P_2 = 0.04,\ P_3 = 0.73,\ P_4 = 0.27 \tag{2}$$

Note, that the rate of overall correct predictions in our experiment is 0.75. In Table 3 we can see the detailed DEA identification results with respect to the type of the sample.

sample (number of non-bankrupt firms in a sample)	10	15	20
NFNB (not on frontier - not bankrupt)	155	183	243
NFB (not on frontier - bankrupt)	0	1	1
FB (on frontier - bankrupt)	20	16	16
FNB (on frontier - not bankrupt)	45	72	97

Table 3: Samples DEA results: The summary

The rates P_1 to P_4 for the samples (see Table 4) indicate that the bankrupt firms remain on the frontier even when the number of non bankrupt firms was increased, i.e. DEA approach seems to perform well in identifying agricultural bankrupt firms. On the other hand, DEA is less powerful in correctly evaluating non-bankrupt agricultural firms. Further we notice that in our experiment the DEA performance is getting slightly worse when increasing the sample volume (i.e. increasing the number of non-bankrupt firms in the sample): while the rates of correctly identified firms are 1 and 0.77 for bankrupt and non-bankrupt firms respectively in the samples with 10 healthy firms, these rates decreases to 0.94, 0.72, and to 0.94, 0.71 for the 15- and 20-healthy-firms-sample, respectively. These results (considering the values of P_1 to P_4 and the rate of overall correct predictions) of our introductory agribusiness experiment correspond to the industry-unspecific result values presented in [12], where these are considered to be promising for further DEA establishment in bankruptcy assessment.

sample (number of non-bankrupt firms in a sample)	10	15	20
P_1	1	0.94	0.94
P_2	0	0.06	0.06
P_3	0.77	0.72	0.71
P_4	0.23	0.28	0.29
overall correct predictions	0.80	0.73	0.73

Table 4: Samples DEA results: The conditional probabilities estimation

4 Conclusion

This contribution aims to provide an entrance validation calculations testing the applicability of a recently suggested DEA bankruptcy assessment technique in agribusiness. According to the results discussed in the previous section, we can conclude that the small test case study provides us with promising numbers. In our future research, further validation calculations using a larger dataset will be done. Using particularly FADN and AMADEUS databases, the dataset can be enlarged to more than 500 agriculture enterprises from central and eastern Europe to provide a base for thorough validation calculations based on evaluating the method performance for sets including increasing number of bankrupted local enterprises and varying in total number of enterprises considered. Further, a detailed comparison with other methods (namely logistic regression) should be provided to arrive to final decision on the effectiveness of suggested method for agribusiness bankruptcy assessment.

Acknowledgements

The research is supported by the grant MSM6215648904 of the Ministry of Education, Youth and Sports of the Czech Republic.

References

[1] Balcombe K., Davidova S., Morrison J.A.: The use of bootstrapped Malmquist indices to reassess productivity change findings: an application to a sample of Polish farms, *Applied Economics*, **40** (16) (2008), 2055–2061.

[2] Bojnec, S., Latruffe, L: Productivity Change in Slovenian Agriculture During the Transition: A Comparison of Production Branches, *Ekonomicky casopis*, **57** (4) (2009), 327–343.

[3] Cesaro, L. et al.: Methodology for analysing competitiveness, efficiency and economy of scale. Use and applications of DEA, *Farm Accountancy Cost Estimation and Policy Analysis of European Agriculture*, Accessible via http://www2.ekon.slu.se/facepa/documents/Deliverable_ D5-1-3_ CUB.pdf

[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] Coelli, T. J. : Recent developments in frontier modelling and efficiency measurement, *Australian Journal of Agricultural Economics*, **39** (3) (1995), 219–245.

[6] Cooper, W.W., Seiford L.M., Tone, K.: *Data Envelopment analysis*. Springer, New York 2007.

[7] Davidova, S., Latruffe, L.: Relationships between Technical Efficiency and Financial Management for Czech Republic Farms, *Journal of Agricultural Economics*, **58** (2) (2007), 269–288.

[8] Emrouznejad, A ., Amin, G. R.: DEA models for ratio data: Convexity consideration, *Applied Mathematical Modelling*, **33** (1) (2009), 486–498.

[9] Keasey, K., Watson, R.: Financial distress prediction models: A review of their usefulness, *British Journal of Mnagement*, **2** (1991), 89–101.

[10] Lindsay, D., Campbell, A.: A chaoss approach to bunkruptcy prediction, *Journal of Applied Business Research*, **12** (1996), 1–9.

[11] Platt, D., Platt, B., Yang, Z.: Probabilistic neural networks in bunkruptcy prediction, *Journal of Business Research*, **44** (1999), 67–74.

[12] Premachandra, I.M., Bhabra G.S., Sueyoshi T.: DEA as a tool for bankruptcy assessment: A comparative study with logistic regression technique, *European Journal of Operational REserach*, **193** (2009), 412–424.

[13] program available at http://www.mathworks.com/matlabcentral/fileexchange/19158-data-envelopment-analysis/content/DEA.m

Mathematical methods in comparative economics

Filip Ježek[1]

Abstract. Comparative economics analyzes and compares the economic systems and processes within these systems, usually within a group of selected countries. The idea is to find common or different characteristics in their development. Subsequently, it can be confronted with an economic level of selected countries. If there were economies that are developed, they would become a kind of "model" for less developed economies. That is how the comparative economics makes it easier to answer questions concerning achieving improvement in the economic area. The comparative economics is a relatively new science. This also concerns the methodological apparatus of comparative economics. For this reason, the article focuses on the description of methods that can be used in comparative economics, especially with regard to mathematical-statistical methods. The article also includes an illustrative example of the use of mathematical methods in comparative economics. The aim is to highlight the potential of comparative economics that can be obtained by using a creative approach in the methodology using mathematical tools.

Keywords: comparison, modeling, corporate income tax

JEL Classification: C51
AMS Classification: 03C98

1 Introduction

Comparative economics as a discipline came into being in 30's in 20th century. The deepening of the differentiation of various parts of the world was the most important incentive. That was a of World War II. Even today there are differences in economic level of countries. Therefore, comparison method is still commonly used. For example Prybyla [7] compares developing countries and analyses impacts of Keynesian revolution, innovations etc. The Importance of using comparative approach is given by the fact that comparative approach can replace the experimentation method, which can not be much used in economic science.

The different set of external conditions (under which the "experiment" is in progress) is simulated by using different characteristics examined in different economies (but the researchers themselves do not set the conditions). Consequently, repeating the "trial" is replaced by comparing several units (i.e. economies).

The analogy of experiment could be comparison of the units in which some intervention took place (e.g. in the form of economic intervention), with units in which this intervention has not taken place. Clasen [1] notes that this quasi-experimental research design suffers from a number of problems (e.g. inability to control effects of other factors).

Comparison can also serve as a method of learning from others, or seeking alternative problem solving. Contrary, comparison can be used to avoid any mistakes. Another important fact is that the comparative method can be used as a specific form of case studies. Comparative research also plays a different role in relation to the theory – it can be useful to validate the theory or to build a new theory. Defining comparative research and situating its value was made by Mills [5].

The comparison focuses not only on comparing the results which are connected to the functioning of economic systems but also to compare the structure of economic systems and compare the mechanisms – how these systems work. It is obvious that mathematical methods are often used just for comparison of results associated with the functioning of economic systems, because the results are easily measurable.

On the other hand there are used qualitative rather than quantitative methods when comparing the structure and mechanism of functioning of economic systems. Here it is necessary to take into account that comparison based on qualitative methods has its limitations - language barriers, different meanings of used words, etc.

An inadequate depth of knowledge of various phenomena and their causes and the lack of knowledge of the overall context can be another limitation for the application of the comparative method based on a qualitative

[1] Silesian University, School of Business Administration, department of economy, Univerzitni náměstí 1934/3,73340 Karvina, Czech Republic. Email:jezek@opf.slu.cz

comparison. This fact points Smelster [10] and emphasizes that a contextual analysis should be done. This often requires an interdisciplinary approach which is connected to another requirement - international cooperation when doing comparison.

For example Geiβler and Mouralová [2] show characteristics of comparative research as follows:

1. The existence of units that are considered to be original entities. These units should be comparable. Entities different in nature or level can not be compared.
2. The cases are mutually related to each other. It is not just a description of the different cases. Therefore, comparison contains generalizations or explanation or interpretation of the differences.

Comparison of the structure and mechanism of functioning of economic systems is complicated not only by the fact that it is necessary to use qualitative methods but also by the fact that the differences in economic systems are difficult to detect and measure. The causes of differences in economic systems can be various objectives (economic objectives may be different in different societies depending on the traditions, laws etc.), the existence of different types of solutions (differences in the of national income distribution), different priorities how to achieve the economic objectives. For these reasons the comparative economics should rather use the quantitative methods. The aim of text below is to show the potential of comparative economics that can be obtained by using a creative approach in the methodology using mathematical tools.

2 Quantitative methods in comparative economics

In comparison where the quantitative methods are used can be seen an important advantage - significant quantity of data. Some difficulties bring the structure of variables and working with aggregated data (for example averages). Aggregation may complicate detection of the effects of economic actions and thus can occur coinciding factors and their effects. The conclusions, evaluation and explanation could be without adequate description unreliable.

The choice of quantitative methodology often leads us to use freely available official data provided by multinational organizations. The data are collected in different countries based on the same methodology (e.g. OECD, Eurostat) The advantage is relatively good comparability and easy accessibility. The disadvantage is superficiality which comes from the nature of the data. In this context Lodge [4] highlights the limitations of such research. The use of international statistics leads to examination of the phenomenon which the state and international organizations "allow" to examine.

In principle, the comparison method is based on classical statistical apparatus. "The comparison especially means that it is compared a set of indicators concerning various countries or average values concerning countries all together. It can be also analyzed the proportion of compared countries to the indicators characterizing the analyzed phenomenon. By Nezval [6] methods of analysis are based on classical statistical apparatus.

Primarily oriented positivist researchers select a higher number of cases (units, entities) because then any conclusions and arguments are more reliable. But it is not possible to go in depth and analysis contains only few explanatory variables. Using quantifiable variables leads to a considerable simplification of the described phenomena. The selection of variables depends on hypotheses and researcher. Therefore, by Ragin [8] it is quite difficult to reveal new and unexpected context. Comparison is often focused on the macro-indicators but the micro-characteristics or internal structure of individual cases could be passed.

Comparative economics generally uses the following methods:

Evaluation - in the case of comparison of economic level for example. Various economic actions in different countries can lead to the same results but on the other hand identical economic actions in different countries could cause different impacts. Comparison therefore monitors and evaluates various actions and their various impacts in various countries.

Prediction - and the subsequent comparison of real (observed, measured) values with expected values (prediction).

Description - descriptive statistical methods are used to describe the obtained data without formulating any conclusions and without verifying any hypotheses or without generalization of any results. Also modeling method belongs here. The model is simplified portrait of reality (avoids from those properties of the observed phenomenon which are not essential to the objective of modeling) and can be useful when trying to understand complex phenomena. Significantly, the models may allow analysis of the problem in less time than in reality. In the model we can change parameters or input variables which allow to obtain a different information. The disadvantage is that the model could simplify reality more than necessary. Another disadvantage is that it can be too much difficult to create the model.

Comparison of seeking a comprehensive solution to the problem of reference should include several steps.

It is necessary to answer the following questions:

- Are any differences in economies?
- How can be the differences explained?
- What are the consequences of these differences?

In the first step are used methods of description and evaluation. The modeling method will be important in the next step. In the third step will be important the modeling method and the method of prediction.

3 Mathematical modeling in comparative economics

As already mentioned, mathematical modeling is one of methods that can be used in comparative economics. Haufler and Wooton [3] focused on foreign direct investment in a region in which the population is asymmetrically distributed between countries and there are some remaining barriers to intra-regional trade. In the following text is shown how to use the method of modeling on the example of comparison of corporate income tax. The reality is that the tax rate on corporate income declined in recent years in most developed countries. As the most frequently mentioned are two reasons - supporting domestic economy and the pursuit of foreign investments. Sedmihradský [9] notes that effect of taxes is often overestimated. But corporate income tax is from the perspective of the state a major source of state revenues but the compare income tax may also affect the inflow of foreign investment and employment. Therefore we have to answer the question what is the highest income tax rate in order to have an investor decided to invest in the country.

The mathematical model described below can determine the tax rate on corporate income in country A in relation to country B so that the country A is attractive for foreign investors and at the same time the country A did not sacrifice more tax revenue than necessary.

This tax rate can be thought as an equilibrium rate because it measures the attractiveness of the country. Then you can also use this model as well as investment decision tool in the matter about which country is the best to place your investments.

Country A and B have symmetric demand curves. In the country A is the demand n-times larger than in country B. The company has to pay a transaction costs per unit when exports. The distribution of goods within only one country means that there are no transaction costs.

The company has four options:

1. import to countries A and B from their home country
2. manufacture in countries A and B
3. place the production to country A
4. place the production to country B

Most interesting is deciding between options 3 and 4. Option 1 is not suitable due to the fact that in this case the transaction costs will be rather high. Option 2 leads to high fixed costs, compared to the transaction costs when producing in country A or country B and importing to the other country. Therefore the company decides whether to place production in country A or B and assuming symmetric transaction costs rather decide for a larger country because in that way the company will serve a larger market. That is the reason why the model also includes the corporate income tax. Sufficiently high taxes on corporate income can lead to decision that company would rather establish a production in a small country. It is necessary to take into account the size of country A which is n-times larger than country B and the existence of transaction costs when importing products from country A to country B. (Assuming symmetric transaction costs means that the transaction costs of importing products from country B to country A are the same when importing from country A to B). Production function is linear and contains labor as variable factor (w) and a capital as fixed factor (K). The amount of wages is different in both countries so that in country A is k-times higher than in country B ($w_A = k * w_B$). Net profit (Π) is the profit after taxation (corporate income tax, $CIT \in [0;1]$).

The consumer demand function:

$$Q_i = \frac{\alpha - p}{\beta} \tag{1}$$

where:

Q = quantity,

p = price,

α and β = parameters determining the elasticity of demand function.

If we got into account that country A is n-times larger than country B we would have:

$$Q_a = \frac{n(\alpha - p)}{\beta} \tag{2}$$

In case that firm decides to produce in country A and export to country B it means that price for consumer in country B will be:

PB = PA + Tr,

where Tr = transaction costs of exporting goods to country B. The transaction costs are symmetric – when exporting form country B to country A, the transaction costs will be again high as Tr.

The function of production is linear and contains variable factor - labour (w) and fixed factor – capital (K). We assume that wages are different in country A is k-times larger than in country B:

wA = k * wB.

The net profit (∏ - profit after taxation by corporate income tax which is linear tax rate in interval [0;1]) can be expressed as follows:

$$\prod A = \{[(P_A - w_A)\,(Q_A(P_A) + Q_B(P_B)] - K\}\ \{1 - TDzpPoA\} \tag{3}$$

When we put all ideas mentioned above into this equation and when we do this in case of country A and B, we can compare both profit functions. When we express the corporate income tax form that, the model gets the following formalized form:

$$CIT_B = 1 - \frac{\left\{\frac{[(n+1)(\alpha - kw_b) - Tr]^2}{4\beta(n+1)} - K\right\}}{\left\{\frac{[(n+1)(\alpha - w_b) - nTr]^2}{4\beta(n+1)} - K\right\}}[1 - CIT_A] \tag{4}$$

where:

CITB = recommended tax rate of corporation income tax in country B (at a level that country B can be attractive to investors)

n = how many times is country A larger than country B

k = how many times are wages higher in country A than in country B

K = amount of invested capital

α, β = parameters to the demand function

Tr = transaction costs of product import

It is obvious that the greatest impact has the tax rate on corporate income in country A (CITA), while the other parameters (especially α, β, w and K) are present in both the numerator and the denominator.

It also shows that the country B can afford a higher corporate income tax when wages in this country (compared to country A) are lower and when corporate income tax in country A is higher.

For comparison of the Czech Republic to the EU countries after inserting data into the model we obtain the following results:[2]

	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
CR vs Belgium	8.9	9.7	9.9	10.3	11.3	11.8	12.8	14.3	15.3	16.0	16.7
CR vs Denmark	8.4	9.0	9.3	10.8	11.1	11.6	12.0	11.2	12.0	12.3	12.8
CR vs Germany	12.0	10.3	10.8	11.7	11.3	11.3	12.4	13.6	11.9	13.0	13.2
CR vs Estonia	28.6	28.7	28.3	28.2	28.1	25.7	24.2	22.7	21.5	21.9	22.1
CR vs Ireland	8.5	7.7	6.3	5.6	5.4	5.1	5.6	6.0	6.7	6.6	6.9
CR vs Spain	21.9	22.0	21.1	22.4	22.9	21.5	19.9	17.7	18.3	18.3	17.8
CR vs Greece	31.0	31.9	30.7	29.3	29.1	30.0	29.8	28.2	27.2	26.4	28.3
CR vs France	15.3	15.5	15.9	18.0	18.6	18.6	19.2	20.3	22.1	22.7	22.7
CR vs Italy	19.7	20.4	23.2	23.8	24.3	24.6	24.9	25.8	21.8	22.0	22.5
CR vs Luxem-bourg	5.8	6.7	6.3	7.3	7.3	8.2	8.2	9.3	10.3	10.4	11.1
CR vs Hungary	20.7	20.3	20.1	20.3	18.3	18.1	18.1	22.2	22.2	22.2	21.8
CR vs Nether-lands	8.3	8.9	9.3	10.8	10.4	9.8	9.2	8.7	9.8	10.4	10.9
CR vs Austria	9.4	11.4	11.9	12.3	11.5	9.1	9.5	10.6	11.7	12.1	12.7
CR vs Poland	28.9	27.2	27.6	27.0	19.3	19.5	19.7	19.5	20.1	20.1	19.9
CR vs Portugal	28.9	29.6	26.2	27.5	23.5	23.1	23.0	22.1	22.9	22.8	22.9
CR vs Slovenia	19.1	19.7	20.0	20.4	20.0	20.1	20.5	19.5	19.0	18.4	17.6
CR vs Slovakia	31.0	31.5	27.6	28.1	21.2	20.8	20.7	20.6	20.5	20.2	20.3
CR vs Finland	12.3	13.5	13.7	14.7	14.2	12.9	13.5	13.7	14.3	14.7	15.2
CR vs Sweden	10.8	11.8	12.5	13.4	13.0	13.9	14.4	14.8	15.8	15.5	16.0
CR vs Britain	5.2	5.7	5.8	6.8	6.5	7.4	8.2	9.3	11.3	11.4	11.7
Min	5.2	5.7	5.8	5.6	5.4	5.1	5.6	6.0	6.7	6.6	6.9
Max	31.0	31.9	30.7	29.3	29.1	30.0	29.8	28.2	27.2	26.4	28.3
Average	16.7	17.1	16.8	17.4	16.4	16.2	16.3	16.5	16.7	16.9	17.1
CIP in ČR	31.0	31.0	31.0	31.0	28.0	26.0	24.0	24.0	21.0	20.0	19.0

Table 1 Recommended tax rate on corporate income for the CR compared with other countries (years 2000 – 2010)

The table shows that reducing tax rate on corporate income in CR was suitable. If you were in CR in 2000 corporate income tax rate was 31% and corporate income tax recommended by the model should be about 17% on average. The CR was not attractive to foreign investors in 2000. The situation has considerably changed and in 2010 the tax rate on corporate income in the CR was closer to the recommended values. We can say that the Czech Republic when compared with other countries is more and more able to compete with other countries.

4 Conclusion

The comparative approach makes it possible to find good and bad examples of practice and allows us to experiment. It gives a basis for predicting the effects of economic actions. Comparative approach forces to explore the broader context of events and contributes to a better understanding of foreign societies and cultures and a deeper understanding of the problem. The comparison is not an objective but rather a method. It uses many other methods including statistical and mathematical methods. Using such methods has its pros and cons. Comparative

[2] Data were obtained from public accessible databases of Eurostat and OECD. This is the tax rate on corporate income (in%) for each country in each year, population in each country and each year (in millions) and information about the amount of average wages in each year and each country. Another data were set as follows: the investor considering an investment (K) of 1 billion, - CZK. Labor costs (w) to produce one unit of output are 20, - CZK. Any transaction costs (Tr) associated with the transport of produce from the country of production to another country is CZK 10, - per unit of production. The demand curve for production has decreasing elasticity -0.2 (the demand function parameters are $\alpha = 100$ and $\beta = 5$).

economics requires a creative approach which also concerns methodology. It is not absolutely necessary to restrict ourselves only to the previously known methods. Searching a new method can extend previously acquired knowledge not only in theoretical but also in practical way. Creating models in comparative economics is an example. The model is simplified portrait of reality but may also allow an analysis of the problem in less time. Searching for suitable mathematical model is therefore a challenge (not only) for the comparative economics.

References

[1] Clasen, J.: Defining comparative social policy. In: *A Handbook of Comparative Social Policy*. Edward Elgar Publishing, 2004.

[2] Geiβler, M. and Mouralová, H.: *Komparativní přístupy při zkoumání veřejných politik.* [online] [cit.2012-04-27]. Available from Internet: http://www.google.cz/url?sa=t&rct=j&q=Komparativn%C3%AD+p%C5%99%C3%ADstupy+p%C5%99i+zkoum%C3%A1n%C3%AD+ve%C5%99ejn%C3%BDch+politik+&source=web&cd=1&ved=0CCQQFjAA&url=http%3A%2F%2Fiss.fsv.cuni.cz%2FISS-138-version1-100224_geissler_mouralova_komparace.doc&ei=RGNgT7zVF6fN4QSFx621Dg&usg=AFQjCNEvZLormEXEtTGrvF4RoWGDffbBfA

[3] Haufler, A. and Wooton, I. *Country Size And Tax competiton for foreign direct investment*. [online] [cit.2012-04-15]. Available from Internet: http://ideas.repec.org/p/lmu/muenec/11120.html

[4] Lodge, M.: Comparative Public Policy. In: *Handbook of Public Policy Analysis: Theory, Politics, and Methods*. (Fisher, R., Miller, G. and Sidny, M.). 2007.

[5] Mills, M. et al.: Comparative Research: Persistent Problems and Promising Solutions. *International Sociology* **21** (2006), 619-631.

[6] Nezval, P.: *Světová ekonomika*. SU-OPF, 2004.

[7] Prybyla J.: *Comparative economics systems.* Ardent Media, New York , 1969.

[8] Ragin, Ch. C.: *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley: University of California Press, 1987.

[9] Sedmihradský, M. *Daňové zatížení jako konkurenční faktor v mezinárodní soutěži.* [online] [cit.2012-05-09]. Available from internet: http://ideas.repec.org/a/fau/fauart/v52y2002i10p550-569.html

[10] Smelser, N. J.: On Comparative Analysis, Interdisciplinarity and Internationalization in Sociology. *International Sociology* **18** (2003), 643-657.

Influence of cyclical development of the most significant foreign-trade partners on small open economy (VAR approach)

Jana Juriová[1]

Abstract. Small open economy is characterized by two basic attributes: significant share of foreign trade and negligible influence on world prices or interest rates. The Slovak Republic (SR) is a typical example of small open economy, as the openness of its economy is approx. 160% according to data of 2011 (defined as the share of total export and import of goods in GDP at current prices).
The main objective of this paper is to identify an influence of foreign factors on the domestic economy. The empirical approach VAR (vector autoregression) is applied for this purpose. Both domestic and foreign economies are represented by basic macroeconomic indicators - gross domestic product, inflation and interest rate and the foreign economy also by the exchange rate and oil price. The most significant foreign partners of the SR are selected according to their maximum contribution to the openness of the Slovak economy. To identify the vector autoregression (VAR), the Cholesky decomposition and structural restrictions are employed. The analysis of reactions of domestic macroeconomic indicators to foreign shocks is performed by means of impulse response functions. The analysis has confirmed that the development of significant foreign-trade partners has a strong influence on the development of the domestic economy.

Keywords: small open economy, model VAR, impulse response functions.

JEL Classification: C50
AMS Classification: 91B84

1 Introduction

After the recent crisis the interest increased in studying macroeconomic fluctuations. In recent global world this issue is especially crucial for open economies. Small open economy is characterized by two main features. The first one is its active participation in international trade and the second is the neglected feedback reaction, i.e. the fact that the small open economy cannot significantly influence the world prices or interest rates.

The basic goal of this paper is to analyse the impact of fluctuations of foreign macroeconomic time series on a small open economy of the Slovak Republic (SR). The SR with its openness[2] of approx. 160% counts among the most open economies of the world. Considering this fact is assumed that the development of foreign environment has a significant impact on the Slovak economy and the foreign disturbances can have a stronger impact than domestic disturbances (in this respect the focus of this paper is mainly on the foreign disturbances). The knowledge of influence intensity of the main foreign factors can help also to reveal the main foreign factors of cyclical development for a small open economy.

In this respect, the empirical evidence is rather scarce in the case of Slovakia. E.g., there is a study of Elbourne and Haan [2] oriented on the monetary policy transmission of Central and Eastern European countries. Relating to the Slovakia's joining of the Euro area in January 2009, the research has been oriented mostly towards the investigation of euro area's influence on the Slovak economy, e.g. in [4]. However, the most important trading partners of Slovakia are also its neighbouring countries which are non-members of euro area. This fact suggests an idea to identify the most important foreign-trade partners of the Slovak Republic and to analyse which foreign shocks caused by the development of the most important trading partners are most significant for a small open economy.

In this paper, modelling of a small open economy uses an empirical VAR approach and first results are gained also from a structural VAR (SVAR) approach. SVAR models are popular in the analysis of sources of business cycle fluctuations, as in Blanchard [1]. According to the literature, the SVAR approach is better suited for small open economies rather than the more traditional identification methods of the VAR, e.g. the Cholesky

[1] VŠB - Technical University of Ostrava, Faculty of Economics, Sokolská třída 33, 701 21 Ostrava, jana.juriova@vsb.cz.
[2] The openness of economy is defined as the share of total import and export of goods in gross domestic product at current prices.

decomposition. For example Elbourne and Haan [2] conclude that the structural VAR yields much better results than the Cholesky ordering, because it can capture more of the salient features of open economies. Cholesky decomposition VAR only allows one direction of contemporaneous causation, which can be a disadvantage in some cases. Both approaches are used for discussion of impulse response functions and identification of the most important business cycle factors.

The paper is organized as follows. Section 2 introduces the theoretical background for the selection of variables and identifies the most important foreign-trade partners of the Slovak economy. Section 3 contains the VAR and SVAR models and discusses their impulse response functions. Section 4 concludes with summary results.

2 Theoretical background and data selection

As the goal is to investigate the impact of foreign business cycle on the domestic business cyclical development, we concentrate on the transmission channels of country-specific shocks. It has been suggested that there are two different channels of international business transmission, as according to Lee [5]. That is, the foreign business cycle has an impact on the domestic economy through its export (through changes in export demand and through changes in terms of trade) and another channel is the financial market. The selected indicators correspond with these two channels. The export demand is contained in the gross domestic product of the trading partners and the changes in terms of trade are influenced by the price level of trading partners. The financial market can be characterized by the development of short term interest rate.

In this model, the Slovak economy is represented by the following basic macroeconomic indicators: real gross domestic product (*y*), inflation (*p*) and short-term interest rate (*r*). The foreign environment is represented by their foreign counterparts of the most important foreign-trade partners (*ys, ps, rs*). The most important foreign-trade partners of SR are identified on the basis of their ratio in the openness of the Slovak economy. The openness of economy is defined as the share of total export and import of goods in GDP at current prices. According to this definition, the following 8 countries of European Union are identified: Germany, Czech Republic, Italy, Austria, Poland, Hungary, France, and United Kingdom. These 8 countries constitute together approx. 63.4% of Slovakia's foreign trade of goods. Only half of these countries are also euro area members (Germany, Italy, Austria, and France).

In the case of open economies also the exchange rate can be included; in this case it is an aggregate exchange rate between Slovakia and the most important trading partners (*e1*). To cover the global economy the exchange rate of Euro to U.S. Dollar (*e2*) and the price of oil (*po*) are also included. The choice of data corresponds to the following economic relationships inspired by the New Keynesian models, e.g. in Horvath [4]: forward-looking pricing rule (1), IS/AD equation (2) and monetary policy rule for setting a short-term interest rate (3).

$$p_{t+1} = a_1 p_t + a_2 p_{t+2} + a_3 y_{t+1} + a_4 (e1_t - e1_{t-1}) + \varepsilon_{t+1} \tag{1}$$

$$y_{t+1} = b_1 y_t + b_2 y_{t+2} - b_3 (r_t - p_{t+1}) + b_4 ys_{t+1} + b_5 e1_{t+1} + \varepsilon_{t+1} \tag{2}$$

$$r_{t+1} = c_1 r_t + c_2 (d_1 y_{t+1} + d_2 \bar{p}_{t+1} + d_3 rs_{t+1} + d_4 ys_{t+1} + d_5 \overline{ps}_{t+1}) + \varepsilon_{t+1} \tag{3}$$

where ε_{t+1} stands for shocks with zero mean value and constant variance. All these equations characterize the relationships between selected variables and all equations can be extended to include also foreign variables like the price of oil and world exchange rate.

3 VAR vs. SVAR model

The starting point of the analysis is a vector autoregression (VAR) model which is free of any economic assumptions. All variables in model VAR are regarded as endogenous. VAR is a system where each variable is regressed on *k* of its own lags and on *k* lags of the other variables. Each equation in the VAR contains the same determining variables and this allows estimating the VAR using OLS method. The reduced form of model VAR in matrix notation is the following:

$$Y_t = A(L) Y_t + u_t \tag{4}$$

where Y_t denotes the *(n x 1)* vector of endogenous variables, *A(L)* denotes the matrix polynomials in the lag operator *L* and u_t is the *(k x 1)* vector of reduced-form errors. However, the estimated model cannot be directly used for the analysis of variables' behaviour in response to the various shocks, as the errors in u_t are correlated with each other. One possible way to solve this is using Cholesky decomposition to identify the underlying orthogonal

shocks. The structural shocks are derived from their reduced-form counterparts through the contemporaneous correlation matrix from the structural moving average representation. It means restrictions such that certain variables have no contemporaneous effect on the others. Cholesky decomposition may be atheoretical, but it implies a strict causal ordering of the variables in the VAR, e.g. the last positioned variable responds contemporaneously to all the other variables, but they do not respond contemporaneously to this variable. For the purpose of better economic interpretation of the impulse responses the economic theory can be included in the model and this leads to the model SVAR - structural vector autoregressions. The structural model SVAR is then given by:

$$\Omega Y_t = B(L)Y_t + e_t \tag{5}$$

The relations between (4) and (5) are the following: $A = \Omega^{-1}B$ and $u_t = \Omega^{-1}e_t$. When we want to express the endogenous variables in Y_t as a function of current and past reduced-form innovations u_t, the moving average representation of (4) is computed:

$$Y_t = C(L)u_t \tag{6}$$

where $C(L) = (I - A(L))^{-1}$. The matrix C can imply pointed restrictions on the system in the form of long-run restrictions.

3.1 Analysis of input data

The analysis is based on quarterly time series from the 1st quarter of 1998 to the 4th quarter of 2011, i.e. 56 observations. All variables in the model are not seasonally adjusted, but transformed into the form of year-on-year growth rates and this transformation removes the seasonality from original data. The data used for the Slovak republic are the following:

- real GDP (*y*) – at constant prices (chain-linked volumes with reference year 2005), mil. EUR, source: Eurostat;
- inflation (*dp*) – quarter-on-quarter changes in harmonized consumer price index (2005=100), source: Eurostat;
- short-term interest rate (*r*), 3-month interbank rate (%), source: OECD.

The most significant trading partners are represented by the following aggregates in the model:

- real GDP (*ys*) – at constant prices (chain-linked volumes with reference year 2005), mil. EUR, source: Eurostat;
- inflation (*dps*) – quarter-on-quarter change in harmonised consumer price index (2005=100), source: Eurostat;
- short-term interest rate (*rs*), 3-months interbank rate (%), source: OECD;
- exchange rate of SR towards the selected trading partners (*e1*) – average monthly nominal exchange rate of SKK/EUR until 31.12.2008 and EUR towards the foreign currency from 1.1.2009, sources: National Bank of Slovakia, Eurostat.

The global economy is characterized by the world exchange rate and price of oil:

- exchange rate EUR/USD (*e2*) – average monthly nominal exchange rate EUR/USD, source: National Bank of Slovakia and
- price of oil (*po*) – price of oil Brent in USD per barrel, source: U.S. Energy Information Administration.

To obtain unbiased estimates of VAR model, the input data are analysed on stationarity by means of Augmented Dickey-Fuller unit root test, all three variants of the test are given in the Table 1.

variable	trend+constant t-stat/sign.	const t-stat/sign.	none t-stat/sign.
y	-3.23(0.09)	**-3.25(0.02)**	-2.24(0.03)
dp	-6.40(0.00)	**-6.46(0.00)**	-6.52(0.00)
r	-4.61(0.00)	-4.59(0.00)	**-4.43(0.00)**
ys	-4.06(0.01)	**-2.90(0.05)**	-2.24(0.03)
dps	-3.38(0.06)	-4.34(0.00)	**-4.38(0.00)**
rs	-1.86(0.66)	-1.60(0.48)	**-1.76(0.07)**
e1	-2.26(0.45)	-2.32(0.17)	**-2.23(0.03)**
e2	-3.42(0.06)	-3.38(0.02)	**-3.23(0.00)**
po	-4.84(0.00)	**-4.90(0.00)**	-4.04(0.00)

Table 1 Testing unit root of variables in levels

The highlighted values are significant for the corresponding variables; trend was insignificant in all the tested time series. According to the results of the test all time series are stationary – I(O) at the significance level of 10%.

3.2 VAR model

VAR model in reduced form (4) is estimated for the vector of endogenous variables $Y_t=(y_t\ dp_t\ r_t\ ys_t\ dps_t\ rs_t\ e1_t\ e2_t\ po_t)$. Regarding the relatively short time series the lag length is set to 2. This lag is verified by the Akaike criterion suitable to use in the cases the model is the best approximation in the information-theoretical sense, as mentioned in [3]. The estimated VAR is stable and its residuals contain no autocorrelation, no heteroskedasticity, however, the normality of residuals is questionable according to the Jarque-Berra test.

To analyse the reaction of domestic economy to foreign shocks the Cholesky decomposition is used. The ordering of variables is proposed on the basis of Granger causality test. Table 2 presents the significance levels at which the hypothesis can be rejected about the uselessness of other endogenous variable to explain the variance of particular variable. With the exception of domestic interest rate the results suggest that all other variables can be explained with this bunch of variables.

y	dp	R	ys	dps	rs	e1	e2	po
0.00	0.01	0.63	0.00	0.02	0.00	0.10	0.06	0.08

Table 2 Results of Granger causality test

The variables are ordered as follows: *(r e1 po e2 dps rs ys dp* y). The impulse response functions follow the effect of shock from one period on innovations of selected time series. The impulses influence the other endogenous variables also indirectly through the dynamic structure of lags in VAR model. The effect of common component is assigned to the first variable in the order.

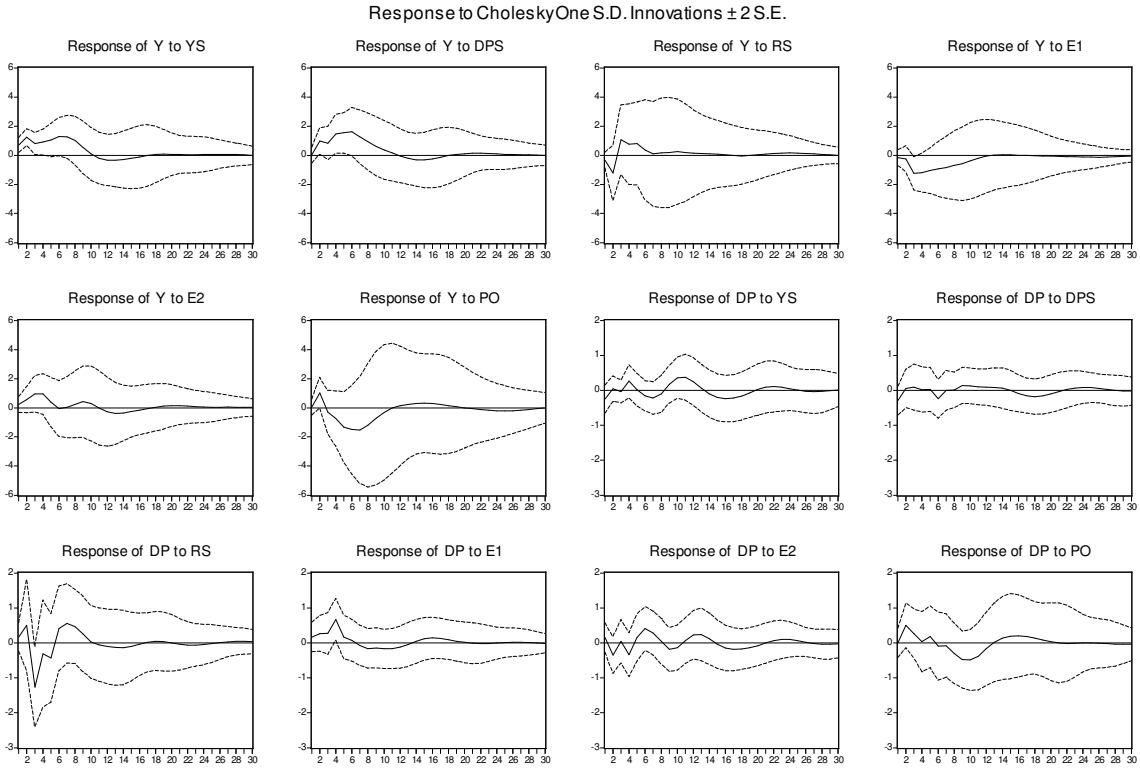


Figure 1 Impulse response functions from Cholesky VAR

Figure 1 contains the impulse response functions for two domestic variables – GDP and inflation (regarding the results of Granger causality test the impact on the interest rate is not analysed). The impact of six following foreign shocks is analysed: foreign demand shock represented by increasing economic activity in trading partners (*ys*), foreign price level (*dps*), foreign interest rate shock (*rs*), two exchange rate shocks - exchange rate of Slovakia vs. trading partners (*e1*) and world exchange rate (*e2*), and supply shock – the development of price of oil at world market (*po*). Each shock simulates a 1% change in standard deviation of innovation. The strong

impact on the domestic economic activity was proved for all analysed shocks; however, the strongest fluctuation was caused by the shock in foreign price level, where the positive peak was achieved after one and half year. The faster reaction of increasing domestic GDP resulted from the foreign demand shock and supply shock – it was recorded already in the second quarter. On the other hand, the negative fluctuations in GDP could be explained with the increased price of oil and increased foreign interest rate. The responses for domestic inflation to foreign shocks show that inflation is not much dependent on the foreign environment. According to the results the highest impact can be assigned to the foreign interest rate where the increase in interest rate caused the decrease in domestic inflation.

3.3 SVAR model

The long-run restrictions imposed on the matrix C are needed for the structural model to be identified. The long-run effect of the proposed shocks on the endogenous variables is given by:

$$\begin{pmatrix} po_t \\ e2_t \\ e1_t \\ rs_t \\ dps_t \\ ys_t \\ r_t \\ dp_t \\ y_t \end{pmatrix} = \begin{pmatrix} C_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & C_{22} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ C_{31} & C_{32} & C_{33} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & C_{43} & C_{44} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{55} & 0 & 0 & 0 & 0 \\ C_{61} & 0 & 0 & C_{64} & C_{65} & C_{66} & 0 & 0 & 0 \\ 0 & C_{72} & 0 & C_{74} & C_{75} & C_{76} & C_{77} & 0 & 0 \\ 0 & 0 & C_{83} & 0 & 0 & C_{86} & C_{87} & C_{88} & 0 \\ C_{91} & C_{92} & C_{93} & 0 & 0 & C_{96} & C_{97} & C_{11} & C_{99} \end{pmatrix} \begin{pmatrix} u_t^{po} \\ u_t^{e2} \\ u_t^{e1} \\ u_t^{rs} \\ u_t^{dps} \\ u_t^{ys} \\ u_t^{r} \\ u_t^{dp} \\ u_t^{y} \end{pmatrix}$$

The restrictions on the foreign variables are imposed according to the correlations of original variables (if the pairwise correlation between variables is below 0.1, the restriction is set to 0) and the restrictions on the domestic variables are set according to the theoretical background – the domestic variables includes the shocks of all variables contained in the equations (1), (2), (3); two insignificant parameters were restricted to 0.

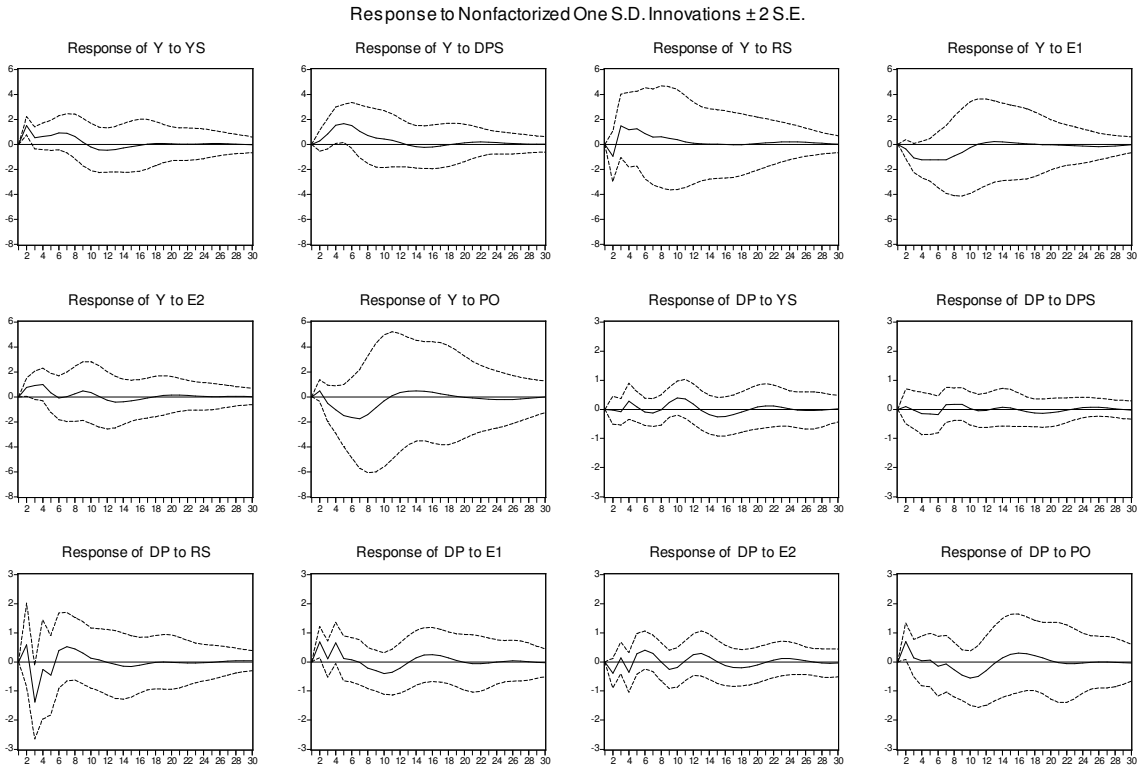


Figure 2 Impulse response functions from SVAR

The model identifies 9 structural shocks, but to be consistent with the assumption of a small open economy, domestic shocks are not allowed to affect the world variables. All parameters of the estimated SVAR are significant and the model is used to analyse the responses of domestic real GDP and inflation to the foreign shocks (Figure 2). At the first glance the results are similar to those from the Cholesky VAR, the highest responses are obtained from the same sources of shocks. Anyway, there are some differences discussed in the last section.

4 Summary results

The analysis of fluctuations of domestic macroeconomic indicators has confirmed the strong impact of foreign shocks coming from the main trading partners, especially on the domestic economic activity. The responses of Slovak indicators are summarized in the Table 3. However, some of the results coming from Cholesky VAR have not to be in compliance with the economic theory, so the model should be adjusted in this respect. The preliminary SVAR model suggests the opposite effects of shocks e.g. in the case of foreign interest rate shock and price of oil shock upon GDP. The domestic inflation responses essentially only to the foreign interest rate shock which causes the decrease in inflation.

shock	peak VAR	timing VAR	average VAR	peak SVAR	timing SVAR	average SVAR
Response of y						
u_t^{ys}	1.25	2	0.27	1.49	2	0.14
u_t^{dps}	1.65	6	0.33	1.66	5	0.30
u_t^{rs}	-1.33	2	0.08	1.49	3	0.25
u_t^{e1}	-1.45	5	-0.34	-1.24	5	-0.28
u_t^{e2}	0.87	3	0.09	1.00	4	0.11
u_t^{po}	1.06	2	-0.10	-1.76	7	-0.25
Response of dp						
u_t^{ys}	0.38	11	0.00	0.39	10	0.02
u_t^{dps}	-0.27	1	-0.01	0.18	9	0.00
u_t^{rs}	-1.26	3	0.00	-1.38	3	-0.01
u_t^{e1}	0.67	4	0.04	0.69	2	0.03
u_t^{e2}	0.41	6	0.01	0.40	6	0.00
u_t^{po}	-0.41	10	-0.02	0.70	2	-0.01

Table 3 Summary of responses for real GDP and inflation

Considering the preliminary version of the presented SVAR model further research should be oriented on the more detailed analysis of economic relations between the macroeconomic variables.

Acknowledgements

This paper was funded by the Student Grant Competition of VŠB - Technical University of Ostrava within the project SP2012/62 “Modelling of business cycle impact on small open economy (VAR approach)”.

References

[1] Blanchard, O. J., and Quah, D. T.: The dynamics effects of aggregate demand and supply disturbances. *The American Economic Review* **79** (1989), 655–673.

[2] Elbourne, A., and de Haan, J.: Modeling monetary policy transmission in acceding countries: vector autoregression versus structural vector autoregression. *Emerging Markets Finance and Trade* **45** (2009), 4-20.

[3] Garrat, A., Lee, K., Pesaran, M. H., and Shin, Y.: *Global and National Macroeconometrics Modelling. A Long Run Structural Approach*. Oxford: University Press, 2006.

[4] Horvath, R., and Rusnak, M.: How important are foreign shocks in a small open economy? The case of Slovakia, *Global Economy Journal* **9** (2009), 1-15.

[5] Lee, H., Huh, J., and Harris, D.: *The relative impact of the US and Japanese business cycles on the Australian economy* **15** (2003), 111-129.

Empirical Estimates in Economic and Financial Problems via Heavy Tails

Vlasta Kaňková[1]

Abstract. Optimization problems depending on a probability measure correspond to many economic and financial applications. Complete knowledge of this measure is necessary to solve exactly these problems. Since this condition is fulfilled only seldom, the problem has to be usually solved on the data basis to obtain statistical estimates of an optimal value and optimal solutions. Great effort has been paid to investigate properties of these estimates; first under assumptions of distribution with thin tails and linear dependence on the probability measure. Recently, it has appeared an investigation in the case of nonlinear dependence on the probability measure and heavy tailed distributions with shape parameter greater two. We focus on the case of the stable and Pareto distributions with shape parameter in the interval (1, 2).

Keywords: Stochastic optimization problems, empirical estimates, thin and heavy tailed distributions, stable distributions, Pareto distribution, Lipschitz property.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Let $\xi\,(:=\xi(\omega)=[\xi_1(\omega),\ldots,\xi_s(\omega)])$ be s–dimensional random vector defined on a probability space $(\Omega,\mathcal{S},P)$; $F(:=F(z),\,z\in R^s)$ the distribution function of ξ; P_F the probability measure corresponding to F. Let, moreover, $g_0(:=g_0(x,\,z))$ be a real–valued function defined on $R^n\times R^s$; $X\subset R^n$ be a nonempty set. If the symbol E_F denotes the operator of mathematical expectation corresponding to F, then a rather general "classical" one–stage stochastic programming problem can be introduced in the form:

Find

$$\varphi(F)=\inf\{\mathsf{E}_F g_0(x,\,\xi)|x\in X\}. \tag{1}$$

In applications very often the "underlying" probability measure P_F has to be replaced by empirical one. Evidently, then the solution is sought w.r.t. the "empirical problem":

Find

$$\varphi(F^N)=\inf\{\mathsf{E}_{F^N} g_0(x,\,\xi)|x\in X\}, \tag{2}$$

where F^N denotes an empirical distribution function determined by a random sample $\{\xi^i\}_{i=1}^N$ corresponding to the distribution function F. If we denote the optimal solutions sets of (1) and (2) by $\mathcal{X}(F)$, $\mathcal{X}(F^N)$, then $\varphi(F^N)$, $\mathcal{X}(F^N)$ are stochastic estimates of $\varphi(F)$, $\mathcal{X}(F)$.

The investigation of these estimates started in 1974 by R. Wets (see [27]). In the same time consistency has been investigated under ergodic assumption in [10]. These papers have been followed by many others (see e.g. [3], [14], [20]). The investigation of the convergence rate started in [11], and followed e.g. by [2], [9], [16], [23], [24],[26]. However these results have been obtained under the assumptions of an linear dependence of objective functions on the probability measure and an "underlying" distribution with "thin" tails. Later have appeared works with underlying distribution functions with heavy tails (see e.g.

[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

[7], [14] and [22]). In this paper we focus on the case of the stable and Pareto distributions with shape parameter from the interval (1, 2).

2 Heavy Tails in Economic and Financial Applications

First, it has been assumed (in the stochastic optimization problems) that the "underlying" probability measure corresponds to the class of distribution functions with thin tails. In particular, the "best" convergence rate has been first proven for the case when an "underlying" objective function has been bounded (see e.g. [11]), later for the case when a finite moment generating function (corresponding to an "underlying" objective function) exits (see e.g. [24]). Consequently, the normal distribution has approximated often the real one. This assumption appeared in the Markowitz model of portfolio selection (see e.g. [4]). However, relatively soon it has been recognized that many data correspond to the distributions for which a finite moment generating function does not exist, it means to the distribution functions with "heavy" tails. This case corresponds often to economic and financial applications.

A relatively good analysis of the "heavy" tailed distribution in economy and finance is presented e.g. in [18]. There is mentioned e.g. the fact that some data about river flow, cotton, exchange rate, returns and so on correspond just to different random parameters with heavy tails distributions. The Weibull distribution corresponds often to lifetime value as well as to problems about wind speed and power, rainfall intensity and so on. Furthermore, it was mentioned in [8] that some data about gold prices, telecommunication, quality control, but also problems about incomes correspond to the lognormal distribution. A relationship between heavy tailed distributions and the stable distributions can be found e.g. in [15], between the stable heavy tailed distributions and the Pareto tails is known and can be found e.g in [15] and [17]. The relationship between tails of stable distributions and Pareto tails distribution is shown in [25].

According to the above recalled facts, it is easy to see that the distributions with "heavy" tails correspond really to many economic and financial data. Consequently, a question arises: how good are empirical estimates corresponding to them. Are these estimates consistent and what is valid about a convergence rate and an asymptotic distribution? Some results about consistency are already known as well as some results about convergence rate. It has been proven that a "best" convergence rate is valid everywhere when all finite absolute moments exist. Weaker results have been obtained generally in dependence on the finite absolute moments existence (for more details see [7]). In this paper we focus on the case of heavy tailed distributions with a shape parameter $\alpha \in (1, 2)$. To this end we employ the stability results obtained by the Wasserstein metric corresponding to $\mathcal{L}_1$ norm.

3 Some Definition and Auxiliary Assertions

Let $F_i, i = 1, \ldots, s$ denote one–dimensional marginal distribution functions corresponding to F; $\mathcal{P}(R^s)$ denote the set of Borel probability measures on $R^s, s \geq 1$; $\mathcal{M}_1(R^s) = \{P \in \mathcal{P}(R^s) : \int\limits_{R^s} \|z\|_s^1 P(dz) < \infty\}$, $\|\cdot\|_s^1$ denote $\mathcal{L}_1$ norm in R^s.

We introduce a system of assumptions:

A.1
- $g_0(x, z)$ is a uniformly continuous function on $X \times R^s$,
- $g_0(x, z)$ is a Lipschitz function of $z \in R^s$ with the Lipschitz constant L (corresponding to the $\mathcal{L}_1$ norm) not depending on x,

A.2
- $\{\xi^i\}_{i=1}^{\infty}$ is independent random sequence corresponding to F,
- F^N is an empirical distribution function 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.

Proposition 1 ([13]). *Let $P_F, P_G \in \mathcal{M}_1(R^s)$, X be a compact set. If A.1 is fulfilled, then*

$$|\varphi(F) - \varphi(G)| \leq L \sum_{i=1}^{s} \int\limits_{-\infty}^{+\infty} |F_i(z_i) - G_i(z_i)| dz_i.$$

Proposition 1 reduces (from the mathematical point of view) s–dimensional case to one–dimensional. Of course, stochastic dependence between components of the random vector is there neglected.

Proposition 2 ([14]). *Let $s = 1$, $t > 0$ and the assumptions A.2, A.3 be fulfilled. If there exists $\beta > 0$, $R := R(N) > 0$ defined on $\mathcal{N}$ such that $R(N) \xrightarrow[N\to\infty]{} \infty$ and, moreover,*

$$\begin{aligned} &N^\beta \int_{-\infty}^{-R(N)} F(z)dz \xrightarrow[N\to\infty]{} 0, \qquad N^\beta \int_{R(N)}^{\infty} [1-F(z)]dz \xrightarrow[N\to\infty]{} 0, \\ &2NF(-R(N)) \xrightarrow[N\to\infty]{} 0, \qquad 2N[1-F(R(N))] \xrightarrow[N\to\infty]{} 0, \\ &(\tfrac{12N^\beta R(N)}{t}+1)\exp\{-2N(\tfrac{t}{12R(N)N^\beta})^2\} \xrightarrow[N\longrightarrow\infty]{} 0, \end{aligned} \tag{3}$$

then

$$P\{\omega:\ N^\beta \int_{-\infty}^{\infty} |F(z)-F^N(z)|dz > t\} \xrightarrow[N\to\infty]{} 0. \tag{4}$$

($\mathcal{N}$ denotes the set of natural numbers.)

Setting $R(N) = N^\gamma,\ \beta > 0,\ \beta+\gamma \in (0,\ \frac{1}{2})$, it follows (under A.2, A.3) from [6] that

$$(\frac{12N^\beta R(N)}{t}+1)\exp\{-2N(\frac{t}{12R(N)N^\beta})^2\} \longrightarrow_{N\longrightarrow\infty} 0.$$

Consequently the validity of the relation (4) depends on the tails behaviour.

Proposition 3 ([7]). *Let $s = 1$, $t > 0$, $r > 0$, the assumptions A.2, A.3 be fulfilled. Let, moreover, ξ be a random variable such that $\mathsf{E}_F|\xi|^r < \infty$. If constants $\beta, \gamma > 0$ fulfil the inequalities $0 < \beta+\gamma < 1/2$, $\gamma > 1/r$, $\beta+(1-r)\gamma < 0$, then*

$$P\{\omega: N^\beta \int_{-\infty}^{\infty} |F(z)-F^N(z)|dz > t\} \longrightarrow_{N\longrightarrow\infty} 0.$$

Analyzing the assertion of Proposition 3 we can obtain for $\beta := \beta(r)$ fulfilling this assertion that

$$\beta \longrightarrow_{r\longrightarrow\infty} 1/2.$$

However, when there exist only first two moments, then there does not exist β (determined by Proposition 3) for which the relation (4) is fulfilled.

Proposition 4 ([1]). *Let $s = 1$, $\{\xi^i\}_{i=1}^N$, $N = 1, 2, \ldots$ be a sequence of independent random values corresponding to a heavy tailed distribution F with the shape parameter $\alpha \in (1, 2)$. Then the sequence*

$$\frac{N}{N^{1/\alpha}} \int_{-\infty}^{\infty} |F^N(z)-F(z)|dz,\ N = 1, \ldots, \tag{5}$$

is stochastically bounded if and only if

$$\sup_{t>0} t^\alpha\, P\{\omega: |\xi| > t\} < \infty. \tag{6}$$

The assertion of Proposition 4 follows from Theorem 2.2 [1]. According to the definition of the stochastically bounded random sequences it holds (under the relation (6)) that

$$\lim_{M\longrightarrow\infty} \sup_N P\{\omega: \frac{N}{N^{1/\alpha}} \int_{-\infty}^{\infty} |F(z)-F^N(z)| > M\} = 0. \tag{7}$$

4 Convergence Rate

Employing the auxiliary assertions of the former part we can obtain new results. However, first we recall some already known assertions.

Theorem 5 ([7]). *Let the assumptions A.1, A.2 and A.3 be fulfilled, $P_F \in \mathcal{M}_1(R^s)$. Then*

$$P\{\omega : |\varphi(F^N) - \varphi(F)| \longrightarrow_{N\longrightarrow\infty} 0\} = 1.$$

Theorem 6 ([7]). *Let the assumptions A.1, A.2 and A.3 be fulfilled, $P_F \in \mathcal{M}_1(R^s)$, $t > 0$. If*

1. *for some $r > 2$ it holds that $\mathsf{E}_{F_i}|\xi|^r < \infty, \quad i = 1, \ldots, s,$*
2. *constants $\beta, \gamma > 0$ fulfil the inequalities $0 < \beta + \gamma < 1/2, \quad \gamma > 1/r, \quad \beta + (1-r)\gamma < 0,$*

then

$$P\{\omega : N^\beta|\varphi(F^N - \varphi(F)| > t\} \longrightarrow_{N\longrightarrow\infty} 0.$$

The last theorem cannot be applied to the stable distributions with the shape parameter $\alpha \in (1, 2)$. Namely, then there does not exist a finite second moment. The following weaker assertion can be proven.

Theorem 7. *Let the assumptions A.1, A.2 and A.3 be fulfilled, $P_F \in \mathcal{M}_1(R^s), M > 0$. If one dimensional components ξ_i, $i = 1, \ldots, s$ of the random vector ξ have the distribution functions F_i with tails parameters $\alpha_i \in (1, 2)$ fulfilling the relations*

$$\sup_{t>0} t^\alpha\, P\{\omega : |\xi_i| > t\} < \infty, \qquad i = 1, 2, \ldots, s,$$

then

$$\lim_{M\longrightarrow\infty} \sup_N P\{\omega : \frac{N}{N^{1/\alpha}}|\varphi(F^N) - \varphi(F)| > M\} = 0 \quad \text{with} \quad \alpha = \min(\alpha_1, \ldots, \alpha_s). \tag{8}$$

Proof. Let $M > 0$, $\alpha \in (1, 2)$. First, it follows from Proposition 1 successively that

$$\begin{aligned}
&\sup_N P\{\omega : \tfrac{N}{N^{1/\alpha}}|\varphi(F^N) - \varphi(F)| > M\} &\leq\\
&\sup_N P\{\omega : \tfrac{N}{N^{1/\alpha}} L \sum_{i=1}^{s} \int_{-\infty}^{\infty} |F^N(z) - F(z)|dz > M\} &\leq\\
&\sum_{i=1}^{s} \sup_N P\{\omega : \tfrac{N}{N^{1/\alpha_i}} \int_{-\infty}^{\infty} |F^N(z) - F(z)|dz > M/Ls\}.
\end{aligned}$$

Consequently, according to Proposition 4 we can obtain

$$\begin{aligned}
&\lim_{M\longrightarrow\infty} \sup_N P\{\omega : \tfrac{N}{N^{1/\alpha}}|\varphi(F^N) - \varphi(F)| > M\} \leq\\
&\sum_{i=1}^{s} \lim_{M'\longrightarrow\infty} \sup_N P\{\omega : \tfrac{N}{N^{1/\alpha_i}} \int_{-\infty}^{\infty} |F^N(z) - F(z)|dz > M'\}, \quad M' = M/Ls.
\end{aligned}$$

Now already we can see that the assertion of Theorem 7 holds. □

Remark 1. Let us assume that the assumptions of Theorem 7 are fulfilled and $\bar\beta \in (0, 1 - 1/\alpha)$, then

$$\lim_{M\longrightarrow\infty} \sup_N P\{\omega : N^{\bar\beta}|\varphi(F^N) - \varphi(F)| > M\} = 0. \tag{9}$$

Setting $\bar\beta := \bar\beta(\alpha) = 1 - \frac{1}{\alpha}$, we can see that $\bar\beta(\alpha)$ is an increasing function of α. Moreover, it holds that

$$\lim_{\alpha\longrightarrow 1^+} \bar\beta(\alpha) = 0, \quad \lim_{\alpha\longrightarrow 2^-} \bar\beta(\alpha) = \frac{1}{2}.$$

5 Special Cases

In this section, first, we recall the definitions of the stable and Pareto univariate distributions.

Definition 1 ([15]). Suppose that $\bar{\xi}^1$, $\bar{\xi}^2$ are i.i.d. non–degenerate random variables. We shall say that $\bar{\xi}^1$ has stable distribution if for any b_1, $b_2 > 0$ there exists $b > 0$ and $a \in R^1$ such that

$$b\bar{\xi}^1 + a =_d b_1\bar{\xi}^1 + b_2\bar{\xi}^2 \quad (=_d \text{ denotes the equality in distribution}).$$

Definition 2 ([8]). A random variable ξ has a Pareto distribution if

$$P\{\omega : \xi > z\} = (\tfrac{C}{z})^\alpha, \qquad f(z) = \alpha C^\alpha z^{-\alpha-1} \qquad \text{for } z \geq C,$$
$$f(z) = 0 \qquad \text{for } z < C, \quad C > 0,\ \alpha > 0 \quad \text{constants.}$$

Let us now analyze these two special cases.

1. If $s = 1$ and ξ has a stable distribution with $1 < \alpha < 2$ (shape parameter), $|\beta| \leq 1$ (skewness parameter) , $\sigma > 0$ (scale parameter) and $\mu \in R$ (location parameter), then according to the fact published e.g. in [25] (page 191) we have

$$\lim_{t\longrightarrow\infty} t^\alpha P\{\xi > t]\} = C_\alpha \frac{1+\beta}{2}\sigma^\alpha, \quad \lim_{t\longrightarrow\infty} t^\alpha P\{\xi < -t\} = C_\alpha \frac{1+\beta}{2}\sigma^\alpha, \tag{10}$$

   where C_α fulfils for $1 < \alpha < 2$ the relation $C_\alpha = \frac{1-\alpha}{\Gamma(2-\alpha)\cos\frac{\pi\alpha}{2}}$. According to the relations (10) and to the fact that the corresponding distribution function belonging to ξ is continuous one we can see that the relation (6) is fulfilled.

2. If $s = 1$ and ξ has a Pareto distribution with a shape parameter α, then

$$t^\alpha\, P\{\omega : |\xi| > t\} = C^\alpha, \quad C^\alpha \quad \text{constant.} \tag{11}$$

   Evidently, the relation (6) is fulfilled in the case of Pareto distribution with $\alpha \in (1, 2)$ too.

Consequently the results of Theorem 7 can be applied to the above mentioned.

Theorem 8. *Let the assumptions A.1, A.2 and A.3 be fulfilled, $P_F \in \mathcal{M}_1(R^s), M > 0$. If one dimensional components ξ_i, $i = 1, \ldots, s$ of the random vector ξ have either Pareto or stable distributions with tails parameters $\alpha_i \in (1, 2)$, $i = 1, \ldots, s$, then*

$$\lim_{M\longrightarrow\infty} \sup_N P\{\omega : \frac{N}{N^{1/\alpha}}|\varphi(F^N) - \varphi(F)| > M\} = 0, \quad \alpha = \min(\alpha_1, \ldots, \alpha_s). \tag{12}$$

6 Conclusion

The paper deals with the investigation of the empirical estimates of the optimal value in the case of one–stage stochastic programming problems with the "underlying" heavy tailed one dimensional marginal distributions and a shape parameters in the interval (1, 2). To this end the results on the stability corresponding to the Wasserstein metric (determined by $\mathcal{L}_1$ norm) have been employed. Evidently, while the "classical" results on the convergence rate have been achieved in the case of finite moments existence, in this case we have proven only that the sequence of difference between "theoretical" optimal value and its empirical approximation corresponds to stochastically bounded random sequence with a convergence rate parameter defined by the value of the shape parameter. However, surely, this assertion generalizes known results.

Acknowledgment This work was supported by the Grant Agency of the Czech Republic under grants P402/10/0956, P402/11/0150 and P402/10/1610.

References

[1] Barrio, E.,. Giné, E., and Matrán, C.: Central limit theorems for the Wassersten distance between the empirical and the true distributions. *Annals of Probability* **27** (1999), 2, 1009–1071.

[2] Dai, L., Chen, C., and Birge, J.H.: Convergence properties of two–stage stochastic programming. *Journal of Optimization Theory Applications* **106** (2000), 489–509.

[3] Dupačová, J. and R. J.-B. Wets, R. J.-B.: Asymptotic behaviour of statistical estimates and optimal solutions of stochastic optimization problems. *Annals of Statistics* **16** (1984), 1517–1549.

[4] Dupačová, J., HurtT, J., and Štěpán, J.: *Stochastic Modelling in Economics and Finance.* Kluwer, London, 2002.

[5] Dupačová, J.: *Portfolio Optimization and Risk Management via Stochastic Programming.* Osaka University Press, Osaka, 2009.

[6] Dvoretski, A., Kiefer, J. and Wolfowitz, J.: Asymptotic minimax character of tesample distribution function and the classical multinominal estimate. *Ann. Math. Statist.* **27** (1956), 642–669.

[7] Houda, M., and Kaňková, V.: Empirical estimates in economic and financial optimization problems. *Bulletin of the Czech Econometric Society* **19** (2012), 29, 50–69.

[8] Johnson, N. L., Kotz, S., and Balakhrishnan, N.: *Continuous Univariate Distributions (Volume 1).* John Wiley & Sons, New York, 1994.

[9] Kaniovski, Y. M., King, A. J., and Wets, R. J.-B.: Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems. *Annals of Oper. Res.* **56** (1995), 189–208.

[10] Kaňková, V: Optimum solution of a stochastic optimization problem with unknown parameters. In: *Trans. 7th Prague Conf. 1974.* Academia, Prague 1977, 239–244.

[11] Kaňková, V.: An approximative solution of stochastic optimization problem. In: *Trans. 8th Prague Conf. 1974.* Academia, Prague 1978, 349–353.

[12] Kaňková, V.: A Note on estimates in stochastic programming. *J Comput. Math.* **56** (1994), 97–112.

[13] Kaňková, V. and Houda M.: Empirical estimates in stochastic programming. In: *Proceedings of Prague Stochastics 2006* (M. Hušková and M. Janžura, eds.), MATFYZPRESS, Prague 2006, 426–436.

[14] V. Kaňková: Empirical estimates in stochastic optimization via distribution tails. *Kybernetika* **46** (2010), 3, 459–471.

[15] Klebanov, L.B.: *Heavy Tailed Distributions.* MATFYZPRESS, Prague, 2003.

[16] Homen–de–Mello, T.: On rates of convergence for stochastic optimization problems under non–i.i.d. sampling. *SIAM J. Optimization* **19** (2008), 524–551.

[17] 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.

[18] Meerschaert, M.M., and Scheffler, H.P.: Portfolio modelling with heavy tailed random vectors. In: *Handbook of Heavy Tailed Distribution in Finance* (S.T. Rachev , ed.), Elsevier, Amsterdam 2003, 595–640.

[19] Omelchenko, V.: Stable distributions and application to finnace. *Diploma Thesis* (supervisor L. Klebanov) Faculty of Mathematics and Physics, Charles University Prague, Prague 2007.

[20] Römisch, W., and Schulz, R.: Stability of solutions for stochastic programs with complete recourse. *Math. of Oper. Res.* **18** (1993), 590–609.

[21] Rachev, S.T., and Mitting, S.: *Stable Paretian Models in Finance (Series in Financial Economics and Quantitative Analysis)* John Wiley & Sons, Chichester, 2000.

[22] Rachev, S.T., and Römisch, W.: Quantitative stability in stochastic programming: the method of probability metrics. *Math. of Oper. Res.* **27** (2002), 4, 792–818.

[23] Shapiro, A., and Xu, H.: Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation. *Optimization* **57** (2008), 395–418.

[24] Shapiro, A., Dentcheva, D., and Ruszczynski, A: *Lectures on Stochastic Programing (Modeling and Theory).* Published by Society for Industrial and Applied Mathematics and Mathematical programming Society, Philadelfia, 2009.

[25] Shiryaev, N.: *Essential in Stochastic Finance.* World Scientific, New Jersey, 1999.

[26] Schulz, R.: Rates of convergence in stochastic programs with complete integer recourse. *SIAM J. Optimization* **6** (1996), 4, 1138–1152.

[27] Wets, R. J. B.: A statistical approach to the solution of stochastic programs with (convex) simple recourse. *Research Report*, University Kentucky, USA 1974.

Usage of analytic hierarchy process for evaluating of regional competitiveness in case of the Czech Republic

Zuzana Kiszová[1], Jan Nevima[2]

Abstract. The contribution solves the problem of alternative access towards evaluating of competitiveness of regions of the Czech Republic. The basic aim of the contribution is due to the method of analytic hierarchy process to define the position of NUTS 2 regions in closed programming period of 2000 – 2006 years. The sense of applying the method will be setting the order of NUTS 2 regions reflecting their competitiveness reached for the year, based on selected criteria. We can obtain the idea of mutual competitive position of these regions by applying the method. The analytic hierarchy process (AHP) is a concrete method of multicriterial decision making method which uses the hierarchy of elements and a pair-wise comparison. The macro-regional indicators will be chosen based on expert estimation regarding to accessibility of relevant statistic data.
Based on the application of the method we can gain detailed view on regional competitiveness of regions by way of quantitative characteristics which can lead to more precise definition of reached competitiveness of NUTS 2 regional units in the EU.

Keywords: competitiveness, region, analytic hierarchy process, European Union

JEL Classification: C49, C61, O18, P48, R11, R58
AMS Classification: 90B50, 90C29, 62H99, 62P20, 91B50, 91B06

1 Introduction

Effectively analyzed competitiveness means to be based on a defined concept of competitiveness. For evaluation of regional competitiveness, we face the problem of the basic concept and definition of competitiveness due to absence of a consistent approach of its definition. Competitiveness has become quite a common term used in many professional and non-specialized publications. Evaluation of the competitiveness issue is not less complicated. In the absence of mainstream views on the assessment of competitiveness, there is sample room for the presentation of individual approaches to its evaluation. In our paper we will examine the possibility of evaluation the competitiveness of the regions of selected Czech regions at NUTS 2 level in terms of analytic hierarchy process. The level of NUTS 2 regions for evaluation of competitiveness seems to be legitimate especially because of the fact that European Commission accents the level of regional units from aims of economic and social cohesion view and realization of structural aid in the EU member states. When making concept of suitable evaluation tools of national and regional competitiveness it is necessary to suggest not only difficult but also simple methods which enable quick evaluation of competitiveness by accessible tools.

2 Concept of regional competitiveness

The concept of competitiveness has quickly spread into the regional level, but the notion of regional competitiveness is also contentious. 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 [2], or geographic concentrations of linked industries [6]. 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. To talk of regional competitiveness would seem to imply that regional economies are like firms [7], or nation-states, and are in competition with one another.

Decomposition of aggregate macroeconomic indicators is the most common used approach at the regional level, as well as comprehensive (mostly descriptive) analysis aimed at identifying the key factors of regional development, productivity [1], [13] and economic growth [11], [12].

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 competitiveness of every

[1] Silesian University in Opava, School of Business Administration in Karviná, Department of Mathematical Methods in Economics, kiszova@opf.slu.cz
[2] Silesian University in Opava, School of Business Administration in Karviná, Department of Economics, nevima@opf.slu.cz

region and major city, and hence the national economy as a whole. Regions play an increasingly important role in the economic development of states.

2.1 Approaches to competitiveness evaluation

Creation of competitiveness evaluation system in terms of the EU is greatly complicated by heterogeneity of countries and regions and also by own approach to the original concept of competitiveness. Evaluation of competitiveness in terms of differences between countries and regions should be measured through complex of economic, social and environmental criteria that can identify imbalanced areas that cause main disparities. Comparing instruments for measuring and evaluation of competitiveness in terms of the EU is not a simply matter. Evaluation of regional competitiveness is determined by the chosen territorial region level, especially in terms of the European Union through the Nomenclature of Territorial Units Statistics (NUTS) – in our paper we apply NUTS 2 level.

Another approach is presented by EU structural indicators evaluation. These indicators are used for the assessment and the attainment of the objectives of the Lisbon Strategy. Specific approach is macro econometric modelling and creation of an econometric panel data model [3]. The approach based on application of specific economic coefficients of efficiency includes two methods of multi-criteria decision making. The first one is the classical *Analytic Hierarchy Process* (AHP) where relevance of criteria's significance is determined by the method of Ivanovic deviation. The second method - *FVK* is a multiplicative version of AHP [4]. Also *DEA* methodology was presented in case of Visegrad four regions. DEA evaluates the efficiency of regions with regard to their ability to transform inputs into outputs [8]. In other words - what results a region can achieve while spending a relatively small number of inputs (resources). This fact is vital for us to perceive the efficiency like a "mirror" of competitiveness [5]. This aspect is also crucial in this paper, where we present analytic hierarchy process to gain more detailed view on competitiveness of regions by way of quantitative characteristics.

2.2 Data base

Data base for evaluation of regional competitiveness in the NUTS 2 regions of the Czech Republic countries is made up of regional data, which were taken from the database of the *OECD iLibrary* – section Statistics - **OECD Regional statistics**. Under regional data have been used time series of 4 indicators (in our case indicators mean "criteria"), annual basis, including: Gross domestic product (GDP), Gross fixed capital formation (GFCF), Gross expenditure on research and development (GERD), Net disposable income of households (NDI). Comparability of data over time was ensured by using time series of the available indicators in purchasing power parity (PPS) per capita in euro currency. The data analysis cover reference period 2000 - 2006.

2.3 Description of entrance criteria for evaluation of competitiveness

GDP was chosen as it is one of the most important macroeconomic aggregate which is simultaneously suitable basic for competitiveness assessment of the country, but also for the regional level, where also NUTS 2 regions belong. It is obviously not always valid that with increasing level of GDP [11] (i.e. increasing efficiency of regions) also the rate of obtained competitiveness/competition advantage grows.

Gross fixed capital formation (GFCF) due to international accounting is a basic part of gross capital (capital investments), in which is also the change of inventories and net acquisition of valuables included. According to ESA 95 methodology GFCF consists of the net assets acquisition minus decrease of fixed assets at residential producers during the time period plus certain increasing towards the value of non-produced assets originated as a consequence of production activity of producers or institutional units. Net fixed capital formation is the difference between gross fixed capital formation and fixed capital consumption. It is estimated in purchase price including costs connected with installment and other costs on transfer of the ownership. Fixed assets are tangible or intangible/invisible assets produced as the output from production process and are used in production process repeatedly or continuously during the one-year period. However, GFCF sense is much broader. It is an index of innovating competitiveness which enables to increase production on modern technical base.

Gross domestic expenditures on research and development (GERD) are sources for further economic growth increasing as stimulation of basic and applied research creates big multiplication effects with long-term efficiency and presumptions for long-term economic growth in economics. R&D is defined as creative work undertaken on a systematic basis in order to increase the stock of knowledge, including human knowledge, culture and society and the use of this stock of knowledge to devise new applications.

Net disposable income (NDI) is the result of current incomes and expenditures, primary and secondary disposal of incomes. It explicitly excludes capital transfers, real profits and loss from possession and consequences of the events as disasters. In contrast to gross disposable income, it does not cover fixed capital consumption.

Disposable income (gross or net) is the source of expenditures on final consumption cover and savings in the sectors: governmental institutions, households and non-profit institutions for households. In sectors of non-financial enterprises and financial institutions is disposable income equal to savings.

3 Analytic hierarchy process

We use multicriteria decision-making method called **analytic hierarchy process** (AHP) to evaluate competitiveness of Czech regions. This method allows including both quantitative and qualitative criteria and is used to determine priorities. Using hierarchies and pairwise comparisons are important attributes of AHP.

Hierarchies allow dividing the problem of evaluation into individual hierarchical levels. The evaluator gets an overview about the problem and its internal relations. Three-level hierarchy is classical example (Figure 1). 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 influence elements on lower level, but not vice versa.

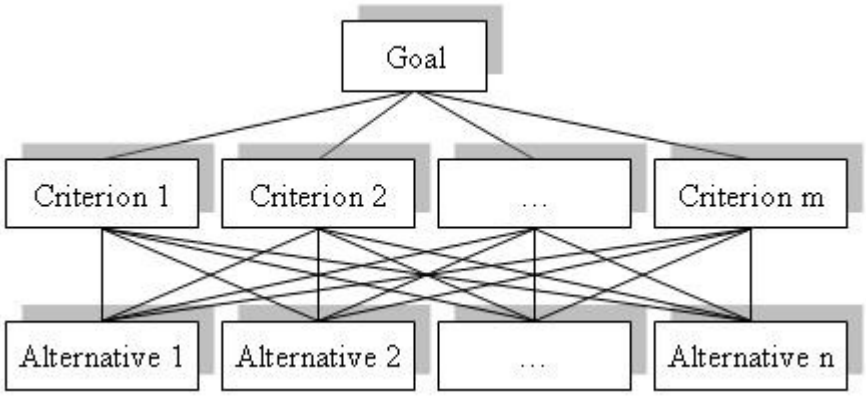


Figure 1 Three-level hierarchic structure

The essence of pairwise comparison is mutual measure of all pairs of considered elements within the same hierarchical level with respect to the level immediately above. We compare criteria among themselves or alternatives with respect to given qualitative criterion. For numerical expression of intensity of relations between compared elements Saaty created nine-point scale [10], see Table 1.

Intensity of importance	Definition
1	Equal importance
2	Weak
3	Moderate Importance
4	Moderate plus
5	Strong Importance
6	Strong plus
7	Very strong Importance
8	Very, very strong
9	Extreme importance

Table 1 Saaty's fundamental scale

Data obtained through pairwise comparisons are inserted into the pairwise comparison matrix A, its entries are signed generally a_{ij}. An n-by-n (square) matrix is created, see Figure 2.

$$\begin{array}{c} \\ \text{element } x_1 \\ \text{element } x_2 \\ \vdots \\ \text{element } x_k \end{array} \begin{array}{c} \begin{array}{cccc} \text{element } x_1 & \text{element } x_2 & \dots & \text{element } x_k \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix} \end{array}$$

Figure 2 General pairwise comparison matrix

Such a matrix is created whenever there is no absolute evaluation of the element with respect to an element from a higher level, i.e. when it is not possible to compare the elements in the given hierarchical level based on their values with respect to an element of the level immediately above. Entries of the pairwise comparison matrix represent estimation of weight ratio of two compared elements of the same hierarchic level (we have to determine these weights through numerical operations). If a_{ij} is an element of pairwise comparison matrix, $a_{ij} \in \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, w_i is wanted weight of the element x_i, w_j is wanted weight of the element x_j for all i and j, we can write:

$$a_{ij} = \frac{w_i}{w_j} \tag{1}$$

$$a_{ji} = \frac{1}{a_{ij}} \tag{2}$$

Formula (2) corresponds to one of the pairwise comparison matrix characteristic – the reciprocity.

We have to compute the eigenvector w corresponding to the maximal eigenvalue λ_{max} of the pairwise comparison matrix A to determine result element priorities of the given matrix. Normalized eigenvector w contents information about result priorities. If I is unit matrix, then:

$$Aw = \lambda_{\max} I \tag{3}$$

Pairwise comparison matrix is square, nonnegative and irreducible. These characteristics ensure existence of maximal eigenvalue λ_{max} and corresponding positive eigenvector [9].

The Wielandt theorem is used to compute the eigenvector, where e is unit vector and c is constant.

$$\lim_{k\to\infty} \frac{A^k e}{\left\|A^k\right\|} = cw, \left\|A^k\right\| \equiv e^T A^k e \tag{4}$$

Through normalization of eigenvectors of individual pairwise comparison matrices we gain weights of criteria with respect to the goal and weights of alternatives with respect to given criterion. The required result, i.e. weights of alternatives with respect to the goal, we obtain through synthesis of these information. If weight of i-th criterion with respect to the goal is w_i and weight of j-th alternative with respect to criterion f_i is $v_j(f_j)$, the overall weight of j-th alternative with respect to the goal is:

$$\sum_{i=1}^{m} w_i v_j(f_i) \tag{5}$$

where $j = 1, 2,\ldots,n$. On the basis of overall weights it is possible to rank evaluated alternatives from the best to the worst. Of course the best alternative gains the highest weight and vice versa.

4 Application

In our case the goal is to assess competitiveness of Czech regions. Alternatives are NUTS 2 Czech regions, i.e. Praha (CZ 01), Stredni Cechy (CZ 02), Jihozapad (CZ 03), Severozapad (CZ 04), Severovychod (CZ 05), Jihovychod (CZ 06), Stredni Morava (CZ 07) and Moravskoslezsko (CZ 08). These alternatives are evaluated by following criteria: gross domestic product (GDP), net disposable income of households (NDI), gross fixed capital formation (GFCF) and gross expenditure on research and development (GERD). All criteria are maximizing.

The pairwise comparison method is applied to determine weights of criteria with respect to the goal. The pairwise comparison matrix of criteria and data base for years 2000 – 2006 are shown in Table 2.

	GDP	*NDI*	*GFCF*	*GERD*
GDP	1	3	5	7
NDI	1/3	1	2	5
GFCF	1/5	1/2	1	2
GERD	1/7	1/5	1/2	1

Year 2000	**GDP**	**GFCF**	**NDI**	**GERD**
CZ 01	26 000	3481.8	8827.3	490.2
CZ 02	12 300	1674.7	6937.6	369.6
CZ 03	12 100	1868.9	6539.4	69.9
CZ 04	10 700	1160.5	6172.6	30.4
CZ 05	11 700	1414.7	6380.2	91.2
CZ 06	11 700	1499.7	6290.2	125.4
CZ 07	10 700	1261.9	6157.4	65.5
CZ 08	10 200	1198.7	6001.7	83.7

Year 2001	**GDP**	**GFCF**	**NDI**	**GERD**
CZ 01	28 700	3883.2	9532.6	499.8
CZ 02	12 900	1959.6	7275.4	377.5
CZ 03	12 800	1837.9	6961.5	72.3
CZ 04	11 000	1648.3	6440.6	29.3
CZ 05	12 300	1354.3	6749.7	94.3
CZ 06	12 700	1491.3	6701.0	119.0
CZ 07	11 200	1712.5	6532.2	67.2
CZ 08	10 800	1521.6	6373.5	80.4

Year 2002	**GDP**	**GFCF**	**NDI**	**GERD**
CZ 01	30 200	5099.0	9707.7	498.5
CZ 02	13 700	1844.1	7591.8	384.5
CZ 03	13 100	1828.4	6968.2	81.1
CZ 04	11 500	1710.3	6401.6	27.0
CZ 05	12 600	1861.9	6833.1	93.6
CZ 06	13 000	1709.0	6778.9	123.3
CZ 07	11 500	1805.0	6648.4	97.3
CZ 08	11 100	1700.3	6420.2	63.2

Year 2003	**GDP**	**GFCF**	**NDI**	**GERD**
CZ 01	31 900	4460.8	10427.5	573.7
CZ 02	14 400	2039.9	8113.6	346.8
CZ 03	13 900	2004.1	7417.0	83.3
CZ 04	12 400	1977.3	6780.1	34.8
CZ 05	13 000	1698.0	7087.6	108.7
CZ 06	13 800	2105.4	7152.4	133.8
CZ 07	12 100	1515.4	7008.0	83.0
CZ 08	11 800	1341.0	6702.0	107.6

Year 2004	**GDP**	**GFCF**	**NDI**	**GERD**
CZ 01	33 400	5505.2	10577.8	613.5
CZ 02	15 400	2260.8	8341.8	342.6
CZ 03	15 100	2061.9	7636.5	90.6
CZ 04	13 200	1640.4	6876.3	29.0
CZ 05	13 800	1801.4	7338.2	123.8
CZ 06	14 600	1857.0	7473.8	147.3
CZ 07	12 900	1654.3	7129.5	80.6
CZ 08	13 300	1430.6	6881.4	94.4

Year 2005	**GDP**	**GFCF**	**NDI**	**GERD**
CZ 01	35 600	5864.4	11225.0	729.9
CZ 02	15 700	2591.6	8823.4	403.8
CZ 03	15 700	2241.1	8175.4	125.8
CZ 04	13 600	1713.9	7362.3	31.9
CZ 05	14 500	1645.2	7939.9	142.6
CZ 06	15 200	2485.5	8012.2	176.3
CZ 07	13 300	1548.6	7641.1	129.6
CZ 08	14 400	1674.4	7559.7	93.6

Year 2006	**GDP**	**GFCF**	**NDI**	**GERD**
CZ 01	38 300	6766.8	12246.7	866.2
CZ 02	17 200	2453.1	9546.0	392.7
CZ 03	16 700	2603.0	8778.6	137.8
CZ 04	14 300	1967.5	7945.8	31.2
CZ 05	15 200	1754.8	8473.9	158.2
CZ 06	16 300	2169.1	8478.9	181.2
CZ 07	14 100	2192.9	8356.8	129.0
CZ 08	15 200	2378.2	7962.1	236.1

Table 2 Pairwise comparison matrix for criteria and data base

MS Excel and complement DAME are used to compute overall priorities. Resulting weights of criteria and positions of particular regions in years 2000 – 2006 are presented in Table 3.

Criterion	**Weight**
GDP	0.5763
NDI	0.2429
GFCF	0.1185
GERD	0.0623

a)

Region/Year	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**
CZ 01	1.	1.	1.	1.	1.	1.	1.
CZ 02	2.	2.	2.	2.	2.	2.	2.
CZ 03	3.	3.	3.	4.	3.	4.	3.
CZ 04	7.	8.	7.	6.	8.	8.	8.
CZ 05	5.	5.	5.	5.	5.	5.	6.
CZ 06	4.	4.	4.	3.	4.	3.	4.
CZ 07	6.	6.	6.	7.	6.	7.	7.
CZ 08	8.	7.	8.	8.	7.	6.	5.

b)

Table 3 a) Weights of criteria, b) rank of regions in years 2000 - 2006

5 Conclusion

In this paper we have dealt with some approach for evaluation of regional competitiveness in the Czech Republic. The final rank of NUTS 2 regions has been presented on the basis of selected macro-regional data using analytic hierarchy process. In Table 3 b) presents final ranks of NUTS 2 regions in the Czech Republic applying 4 criteria (presented by GDP, GFCF, NDI and GERD) and 8 alternatives (presented by 8 NUTS 2 regions). For example the rank of extreme cases – CZ 01 and CZ 02 region remain unchanged. For example region CZ 08 indicates change in rank during evaluated period (8th position in 2000 and 5th position in 2006). On the other hand, each technique is specific so we can not say that some leads to more (or less) credible result than the others. Table 3 a) also shows weights of 4 criteria. The highest impact has GDP – the most importance macroeconomic aggregate. GERD has the lowest impact to final rank of regions. Consequently, our approach presented here could be considered as a suitable alternative for the evaluation of regional competitiveness not only in the Czech Republic. We intend to apply approach of AHP in case of Visegrad four in future.

Acknowledgements

This paper was created in the students' grant project "Seeking Factors and Barriers of the Competitiveness by Using of Selected Quantitative Methods". Project registration number is SGS/1/2012.

References

[1] Hančlová, J.: Panel Modelling of Globalization on Government Expenditures for the Selected New EU Countries. In: *Liberec Economics Forum 2011*. Technical university of Liberec, Faculty of economics, Liberec, 2011, 161-170.

[2] Mazurek, J.: Comparison of impacts of the financial crisis on the Czech Republic regions by cluster analysis. In: *Proceedings of the 13th International conference on finance and banking*. Silesian University, School of Business Administration, Karviná, 2011, 396-404.

[3] Nevima, J., and Melecký L.: Regional competitiveness evaluation of Visegrad Four countries through econometric panel data model. In: *Liberec Economics Forum 2011*. Technical university of Liberec, Faculty of economics, Liberec, 2011, 348-361.

[4] Nevima, J., and Ramík, J.: Application of multicriteria decision making for evaluation of regional competitiveness. In: *Mathematical Methods in Economics 2009*. Czech University of Life Sciences Prague, Kostelec nad Černými lesy, 2009, 239-244.

[5] Nevima, J., and Ramík, J.: Application of DEA for evaluation of regional efficiency of EU regions. In: *Mathematical Methods in Economics 2010*. University of South Bohemia, České Budějovice, 2010, 477-482.

[6] Porter, M. E.: The Economic performance of regions. *Regional Studies* **37** (2003), 549-578.

[7] Pražák, P.: On a dynamic model of revenue maximizing firm. In: *Mathematical Methods in Economics 2010*, University of South Bohemia, České Budějovice, 2010, 542-547.

[8] Ramík, J., and Hančlová, J.: Multicriteria methods for evaluating competitiveness of regions in V4 countries. In: *Trzaskalik. T. Multiple Criteria Decision Making'12*. Katowice: The Karol Adamiecki University of Economics, 2012.

[9] Ramík, J., Perzina, R.: *Modern methods of evaluation and decision making* (*Moderní metody hodnocení a rozhodování)*. OPF SU, Karviná, 2008.

[10] Saaty, T. L.: *The analytic hierarchy process: Planning, priority setting, resource allocation*. McGraw Hill, New York, 1980.

[11] Tuleja, P., and Tvrdoň, M.: The Czech labour market after the crisis of a real economy: negative development or return to steady-state? *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* **LIX** (2011), 477-488.

[12] Tvrdoň, M., and Verner, T.: Examining the relationship between economic performance and unemployment: the case of Visegrad countries. In: *Mathematical Methods in Economics 2011*, University of Economics, Praha, 2011, 733-738.

[13] Viturka, M.: Competitiveness of regions and possibilities of its evaluation (Konkurenceschopnost regionů a možnosti jejího hodnocení). *Politická ekonomie* **5** (2007), 637-657.

VAR model with current-optimal leading indicators

Miroslav Kľúčik[1]

Abstract. Mutual dependence of business cycles is not easy tractable in the tangle of interconnected economic relationships. The small and open Slovak economy is dependent on the performance of its main trading partners and thus faces the fluctuations of the global economy. For managers and economic policy makers it is crucial to interpret different signals from various segments of the economy for their decisions about future activities. That information is currently available in immense quantity. An artificial intelligence tool – genetic programming – can exploit the huge amount of available data and find patterns of associations between the Slovak economy and foreign economies. Symbolic regression via genetic programming is used. The individual time series and their combinations are optimized for best fit and optimal lead against the Slovak economy. Such leading indicators are used for forecasting of the Slovak economy in the structure of a VAR model. The forecasting ability is tested and compared to a proxy AR model. The forecasts of the VAR model using current-optimal leading indicators show advanced quality compared to the proxy model.

Keywords: business cycle, leading indicators, VAR model, genetic programming, symbolic regression.

JEL Classification: C32, C53, C63, E32
AMS Classification: 60G50, 68T05

1 Introduction

Fluctuations of a small and open economy are mainly driven by the foreign markets, specifically, by its main trading partners. The potential information useful for economic decisions of market participants is therefore available in large quantities considered the macroeconomic data of each trading partner. The main aim is to exploit the immense amount of information and to extract patterns of associations between the business cycle of a small and open economy of Slovakia and other economies.

Business cycle analysis and forecasts can be based on model and non-model approach. The simple non-model approach, as e.g. the OECD methodology for leading indicators construction [14], fulfils mainly the task of qualitative forecast of economic activity. On the other side, model approaches enable to make quantitative forecasts using the leading indicators. The observed non-linearity in business cycles, e.g. [6], [15], gives particular attention to nonlinear models such as regime switching models, models using neural networks or evolutionary computations.

The aim of this paper is to transform the original leading indicator of the Slovak economy as in [7], intended only for qualitative forecasts, to a model approach based on leading indicators and also to improve the adaptability of the leading indicator in relation to the non-linear character of business cycles.

Basic modelling approaches for business cycles are the unobserved variables based models (dynamic factor models), regime switching models or the classic VAR model. A VAR model using leading indicators was constructed by Mendez et al. [12], Cubbada, Hecq [3], Fichtner et al. [5] or Savin and Winker [16].

For the composite leading indicator construction, i.e. for improving its ability to adapt to nonlinear characteristics of business cycle (e.g. different development of cycle in recessions and expansions), the most suitable is the application of heuristics. This is because the space of potential solutions is very large given the more than 20 thousand of time series available in the database and beforehand unknown character of the relationship between the cycles. The group of evolution computation techniques can be regarded as a very effective data mining method for large databases, i.e. genetic algorithms, genetic programming (GP) and other evolutionary computation methods. GP allows the most abstract handling of problems while using small computer programs as individuals in the evolution process in search for the global optimum.

Although GP is predominantly applied in areas of financial markets (high frequency data) some attention was paid also to macroeconomic time series modelling, e.g. in forecasting of Gross Domestic Product (GDP) [11],

[1] VSB – Technical University Ostrava, Faculty of Economics, Department of Economics, Sokolská třída 33, 701 21 Ostrava, Czech Republic, miroslav.klucik@vsb.cz.

[17] or other time series [10], country risk early warning system as in [2]. Genetic algorithms have been used in optimizing regression coefficients for private housing demand leading indicators models [13] or business cycles indicators selection based on genetic algorithm based clustering method [18]. Kotanchek et al. [8] uses genetic programming for detecting models and outliers in large public data sets and Kronberger et al. [9] identifies variable interactions using GP symbolic regression and macroeconomic time series.

The outline of this paper is the following: the 2nd chapter consists of introducing the database of time series and a basic genetic programming model allowing the search for the best matching composite leading indicator for Slovak economy (identification of leading indicators, comparison with a classical average composite indicator). In the 3rd chapter the composite leading indicators are used for constructing a simple VAR model. This is used for forecasting of the Slovak GDP retrospectively for the years 2008-2011. A simple autoregressive model is taken as a proxy for comparison of forecasts. The main findings are concluded in the 4th chapter.

2 Search for leading indicators

The database of time series gathered by the author consists of basic macroeconomic indicators of the European economies and other trade partner of Slovakia (USA, China, South Korea etc.). The analysis is based on quarterly data from the sources of Eurostat, OECD and IMF. The data cover the areas of the whole economy – national accounts data, different branches of real economy (industry, construction, retail trade, services), consumer and business tendency surveys, financial markets data (equity and commodity indexes, exchange rates, short-term and long-term interest), employment, consumer and producer prices, foreign trade data etc.

Together the database comprises over 21 thousand time series. This amount is reduced to over 9 thousand time series by requiring full sample from 1998 to 2011. To avoid unnecessary inaccuracies connected with seasonal adjustment, all the data are transformed to year-over-year changes. Exceptions are the data from business and consumer surveys and time series of balances. These are understood as a deviation from long-term growth, i.e. they are comparable to the year-over-year changes of other time series. However, these time series need seasonal adjustment, therefore both versions of the data are used in the analysis (seasonally adjusted and not adjusted).

2.1 Basic GP model

Genetic programming is a natural selection based algorithm successfully tested mainly for searching large spaces of potential solutions. The aim is to apply symbolic regression of GP for finding the best leading indicators for Slovakia.

The procedure of GP is the following: consider a population where each single individual represents a solution of a problem. In a process of natural selection the best individuals (best solutions) are more likely to be successful in passing the good genes onto the next generation. With the continuing selection procedure new generations are arising with better solutions. Generally, each individual tries to adapt to the current environment through crossover with other individuals, i.e. each solution attempts to get closer to the global solution of the modelled problem. Individuals are encoded in trees (tree representation of individuals). The trees exchange their branches or leafs (crossover) and can mutate.

The search for the best model is based on symbolic regression, where the GDP of Slovakia is the dependent variable and other time series from the database are the potential explanatory variables. The explanatory variables are leafs of the tree, together with constants. The roots of the branches are the terminals – functions (+, -, *, /). Best individuals are chosen according to fitness of individual solutions, which is in the case of symbolic regression a difference measure between the estimated model and true values. Symbolic regression contrary to the classical regression does not assume a relationship between the dependent and explanatory variables in advance. The resulting relationship is mostly nonlinear.

This procedure is written in the EViews language by the author of this paper. The basic structure of the program is the following:

Random population
For (number of generations)
Evaluation of individuals (fitness measure)
Crossover (crossover probability)
Mutation (mutation probability)
⇨ *New individuals – new generation*
Next

The GP parameters for controlling the process of evolution are the following: number of generations, number of initial population, number of constants, selection method, tournament size, number of elites, probability of crossover, probability of mutation and other restrictions (maximum depth of individuals, constant population) etc. The optimal parameters are set according to a sensitivity analysis.

The choice of optimal parameters is based on 10 preliminary runs of sensitivity analysis for each variation of the parameter. The optimal parameter is chosen according to the best average fitness measure of the individuals in last generation. In this case the RMSE measure is used (root mean square error). The final parameters are the following: initial population number – 1000 individuals, number of constants – 1000, tournament size – 29, number of elites in tournament – 1, probability of crossover – 0.9, probability of mutation – 0.01[2].

No considerable progress in fitness has been recorded after the 50th generation (Figure 1); therefore this value is set as the termination criterion.

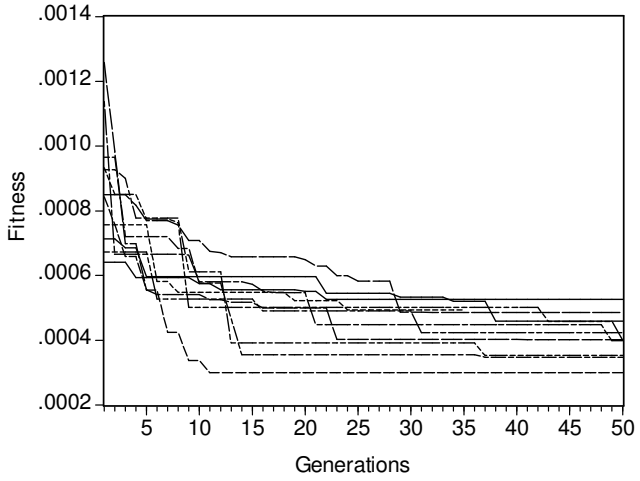


Figure 1 Fitness progress

2.2 Results and comparison

Total of 50 runs have been executed on the GP symbolic regression with maximum of 50 generations in each run. It means there are 50 best individuals/solutions/models available. Except these competing models we can evaluate best individuals in each tournament (29 in each run). The results are mostly nonlinear solutions, but relatively simple, which is acceptable in view of the problems of over-fitting. Evaluating the structure of the best individuals the following variables are mostly repeated in the trees structure as explanatory variables for the Slovak GDP: final consumption expenditure of general government in Euro area (leading 3 quarters), deflated turnover index of Euro area (+ 1 quarter), final consumption expenditure of Austria (+ 1 quarter), GDP of Czech Republic (+ 2 quarters), final consumption expenditure of Czech Republic (+ 1 quarter), gross capital formation of the European Union (EU - + 1 quarter), external balance of goods of Germany (+ 1 quarter), competitive position over the past 3 months EU companies (+ 1 quarter) etc. Some of the variables were not taken into account due to their potential spuriousness, e.g. the GDP of Norway, Denmark etc. The relationship between Slovakia and Norway or Denmark is not direct, but could be regarded as indirect (e.g. through the relationships with Germany). This assumption needs rather further analysis and therefore it is left out from this work.

As an example the formula of the best individual is the following[3]:

$$GDP_{SK} = \big((\_NAB111DE(+2))(\_NAB112EU27(+2))(\_NAB111DE(+1))(\_NA0B1GMCZ(+1))\big) + E_t \qquad (1)$$

To compare the performance of the above GP best individual, a simple composite leading indicator can be used as a proxy. The most associated time series from the database of over 9 thousand time series are chosen according to cross correlation with the GDP of Slovakia. Omitted are again time series with potential threat of spuriousness. Finally, the composite leading indicator used as a proxy is computed as a simple average of the

[2] Calculations performed by the author are provided in the documentation available by request.

[3] GDP_{SK} – GDP of Slovakia in constant prices, _NAB111DE(+2) – external balance of goods of Germany (leading 2 quarters), _NAB112EU27(+2) – external balance of services of EU27 (+ 2 quarters), _NAB111DE(+1) – external balance of goods of Germany (+ 1 quarter), _NA0B1GMCZ(+1) – GDP of Czech Republic (+ 1 quarter).

following five time series: retail confidence indicator of EU27 (leading 1 quarter), Dow Jones Euro Stoxx Basic Materials index Euro area (changing composition, + 1 quarter), trend of activity compared with preceding months in UK (+ 1 quarter), expected business situation in EU27 (+ 1 quarter), balance of goods of Germany (+ 1 quarter). The comparison is depicted in Figure 2 (equalized phases – time shifted series).

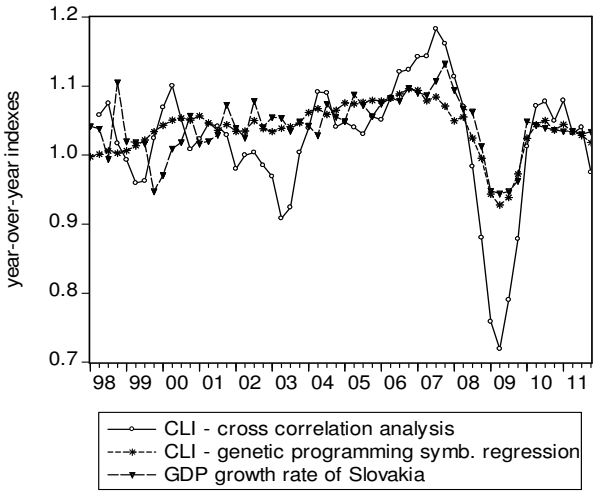


Figure 2 Comparison of composite leading indicators

Clearly, the composite leading indicator from the GP symbolic regression performs better than the leading indicator from cross-correlation analysis (Figure 2), regarding the size of fluctuations and generally, the fitness between the time series.

3 VAR model forecast

Enabling the use of leading indicator for forecasts, the competing 50 composite leading indicators from GP runs are taken preliminary for the VAR model construction. A simple VAR model of the following form is regarded:

$$Y_t = a_0 + A_1 Y_{t-1} + ... + A_p Y_{t-p} + B_1 X_{t-1} + ... + B_q X_{t-q} + E_t \tag{2}$$

The matrices of parameters of endogenous variables (of vector Y_p) are denoted A_p and of exogenous variables (of vector X_p) as B_q, where p, q are lags. E_t is a vector of random disturbances and a_0 the constant.This form of VAR model is called VARX model, while it includes also exogenous variables. Initially, the composite leading indicator is regarded as an exogenous variable considering only one way dependence relationship between the small Slovak economy and outside economies.

The primary condition of endogenous and exogenous variables for entering the VAR model is the stationarity of the time series. This condition is tested for the 50 leading indicators via the ADF test. All time series are stationary at 10% significance level. This was to be expected due to the year-over-year transformation of the time series (removed trend).

3.1 Model selection

The VAR model is constructed gradually according to the basic criteria deciding about the usability and quality of the model: stationarity of the VAR model, exogeneity test of endogenous variables, test of lag exclusions and residual test.

In the first step the optimum number of lags is determined using the Akaike information criterion. Maximum 6 lags are taken for the test arbitrary. The model of the lag order with lowest AIC value is chosen and consequently, non-significant lags are excluded according to the Wald statistics (5% significance level). All 50 models are estimated using the above mentioned properties. The stationarity of each model is judged following the value of roots of autoregressive polynomials, which must lie inside the unit circle. The test indicates stationarity of all 50 models. Initially the composite leading indicator was regarded as exogenous variable; this assumption is confirmed by the two-way Granger causality test. As expected the GP composite leading indicator indicates only one-way relationship according to the test – from the composite indicator to the Slovak GDP, this is confirmative for all 50 GP competing indicators. Lastly, the presence of autocorrelation in the residuals is tested via the Breusch-Godfrey LM test and the normality of residuals. All of the VARX models do not violate any of these conditions (non-presence of autocorrelation and normality of residuals) at a 5 % significance level. The model

with the best individual is chosen for the forecasts. The model is estimated with 5 lags, with 2nd a 3rd lag of the endogenous variable excluded. The exogenous variable (composite leading indicator) is presented in the same form as in (1).

3.2 Forecasts

The predictive ability of the VARX model is tested on the sample 2008-2011, covering the beginning of the current financial/economic crisis in Slovakia (3rd quarter of 2008) and the slow economic recovery since 2010. The robustness of the model is questionable taking into account the short quarterly time series and properties of the model in adjusted sample, but nevertheless the forecasting ability is tested in two steps for each quarter - estimation of the VARX model, one-quarter forecast, two-quarter forecast, prolongation of the sample by one quarter and anew estimation of the model with forecasts for one and two quarters ahead, and so on. The insignificant parameters of the model have been excluded stepwise (due to over-fitting) from the estimated GDP equation of the model following [1] and [4].

An AR model is used as a proxy model with 5 lags (same number of lags as the VARX model). We can also compare the VARX model with GP indicator with a VARX model containing the composite indicator from correlation analysis (with evaluated properties as the previous VARX model). For evaluation of the forecast the Theil's U is used (TU), which is the share of RMSFE (root mean squared forecast error) of the VARX model (e_t) on the forecast error of the AR proxy model (u_t):

$$TU = \sqrt{\sum_{t=T_1}^{T_2} e_t^2} \Big/ \sqrt{\sum_{t=T_1}^{T_2} u_t^2} \tag{3}$$

$$RMSFE = \sqrt{\frac{1}{T_2 - T_1 + 1} \sum_{t=T_1}^{T_2} e_t^2} \tag{4}$$

The T_1 and T_2 is the first and the last forecasting period, e_t is the difference between the true value and forecasted value of the GDP growth rate. The forecast with the lowest RMSFE is considered the best.

In Figure 3 the graphical comparison of VARX model forecast and AR forecast is shown.

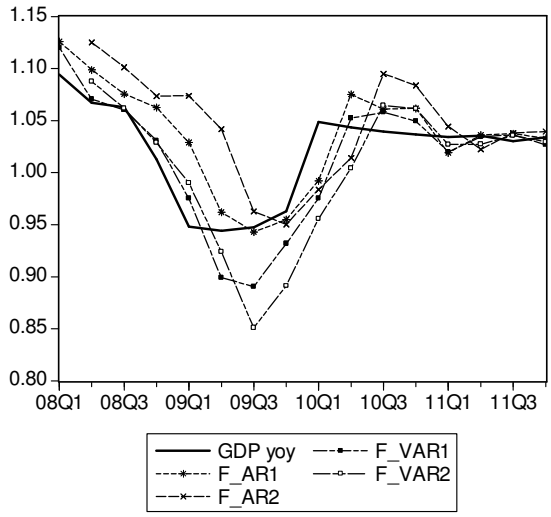


Figure 3 Forecasts comparison

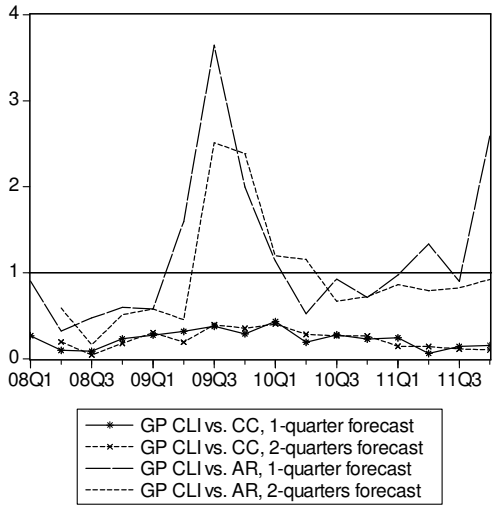


Figure 4 Theil's U – evaluation of forecasts

The VARX model based on the GP composite leading indicator is outdoing the AR model forecasts in 10 out of 16 cases for the one-quarter forecasts and in 11 out of 15 cases for the two-quarter forecast. Figure 3 shows the performance of both forecasts. Also, surprisingly, the GP based VARX model has outperformed all the forecast from the VARX model with composite leading indicator based on the correlation analysis – CLI CC (from Figure 2). The results of the Theil's U are given in Figure 4 above. Theil's U below 1 denotes better forecast of the GP based VARX model. The VARX model shows lower forecast ability during the period of GDP growth after the business cycle turning point in 2009 and in the period of slow GDP growth in 2011.

4 Conclusions

This work attempts to quantify the forecasts of composite leading indicator using a VAR model. Additionally, the forecast is improved by using genetic programming symbolic regression for the composite leading indicator construction. VAR model with exogenous variable representing the leading indicator based on genetic programming clearly outperforms a VAR model with simple average composite indicator and in most cases also an AR proxy model. This proves the possibility of improving the forecasts based on linear relationships between leading indicators by simple nonlinear models. These models have better potential to adapt to the fluctuations of business cycles.

Acknowledgements

This work was financially supported by the Student Grant Competition of Technical University of Ostrava, Faculty of Economics, within the project SP2012/62 "Modelling the impact of cyclical development on a small open economy (VAR approach)".

References

[1] Alles, M. G., Kogan, A., Vasarhelyi, M. A. and Wu, J.: *Continuous data level auditing: Business process based analytic procedures in an unconstrained date environment*. Department of accounting & information systems, Rutgers Business School, 2006.

[2] Apoteker, T. and Barthelemy, S.: *Genetic Algorithms and Financial Crises in Emerging Markets*. TAC Financial, Saint Hilaire des Landes, 2000.

[3] Cubadda, G. and Hecq, A.: The Role of Common Cyclical Features for Coincident and Leading Indexes Building. *Economics & Statistics Discussion Papers*. University of Molise, Dept. SEGeS, 2003.

[4] Di Fonzo, T. and Marini, M.: *Benchmarking a system of time series: Denton's movement preservation principle vs. data based procedure.* University of Padova, 2005.

[5] Fichtner, F., Rüffer, R., and Schnatz, B.: Leading indicators in a globalised world. *Working Paper Series 1125*, European Central Bank, 2009.

[6] Hamilton, J. D.: A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica* **57** (1989) 2, 357–384.

[7] Kľúčik, M.: *Composite Indicators for Small Open Economy: The Case of Slovakia*. 30th CIRET conference, New York, United States, 2010. Retrieved from: https://www.ciret.org/conferences/newyork_2010/papers/upload/p_54-548193.pdf

[8] Kotanchek, M. E, Vladislavleva, E. Y., and Smits, G. F.: Symbolic Regression Via Genetic Programming as a Discovery Engine: Insights on Outliers and Prototypes. In: *Genetic Programming Theory and Practice VII, Genetic and Evolutionary Computation* (Riolo, R. et al., eds.), Vol. 8, 55-72 Springer Science+Business Media, LLC, 2010.

[9] Kronberger, G., Fink, S., Kommenda, M., and Affenzeller, M.: Macro-economic Time Series Modeling and Interaction Networks. In: *EvoApplications* (Di Chio, C. et al., eds.), Part II, LNCS 6625, 2011, 101–110.

[10] Lee ,Y. S. and Tong, L. I.: Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming. *Knowledge-Based Systems* **24** (2011) 1, 66-72.

[11] Li, M., Liu, G., and Zhao, Y.: Forecasting GDP growth using genetic programming. *Proceedings of the Third International Conference on Natural Computation* **4**, IEEE Computer Society Washington, DC (2007), 393-397.

[12] Mendez, G. C., Kapetanios, G., and Weale, M. R.: The Forecasting Performance of the OECD Composite Leading Indicators for France, Germany, Italy and the UK. *NIESR Discussion Papers 155*, National Institute of Economic and Social Research, 1999.

[13] Ng, S. T., Skitmore, M., and Wong, K. F.: Using genetic algorithms and linear regression analysis for private housing demand forecast. *Building and Environment* **43** (2008) 6, 1171-1184.

[14] OECD: *OECD System of Composite Leading Indicators,* 2008. Retrieved from: http://www.oecd.org/dataoecd/26/39/41629509.pdf

[15] Paap, R., Segers, and R. Dijk, D.J.C. van: Do Leading Indicators lead Peaks more than Troughs? *Journal of Business & Economics Statistics* **27** (2009), 528–543.

[16] Savin, I. and Winker, P.: Heuristic model selection for leading indicators in Russia and Germany. *Working Papers 046*, COMISEF, 2011.

[17] Wagner, N. and Brauer, J.: Using dynamic forecasting genetic programming (DFGP) to forecast United States gross domestic product (US GDP) with military expenditure as an explanatory. *Defence and Peace Economics*, **18** (2007) 5, 451-466.

[18] Zhang, D., Yu, L., Wang, S., and Song, Y.: A novel PPGA-based clustering analysis method for business cycle indicator selection. *Frontiers of Computer Science in China* **3** (2009) 2, 217-225.

Short-term and long-term growth effects of exchange rate adjustment

Evžen Kočenda[1], Mathilde Maurel[2], Gunther Schnabl[3]

Abstract. During the European sovereign debt crisis the discussion concerning the pros and cons of exchange rate adjustment in the face of asymmetric shocks has been revived. Whereas one side has recommended (in the spirit of Keynes) the exit from the euro area to regain rapidly international competitiveness, exchange rate stability cum structural reforms have been argued (in the spirit of Schumpeter) to be a beneficial long-term strategy towards the reanimation of a robust growth performance. Previous literature has estimated the average growth performance of countries with different degrees of exchange rate volatility. We augment this literature by econometrically separating between the short-term and long-term growth effects of exchange rate volatility based on a panel-cointegration framework for a sample of 60 countries clustered in five country groups. The estimations show that countries with a low degree of exchange rate volatility exhibit a higher long-term growth performance, whereas over the short-run exchange rate flexibility provides some benefits.

Keywords: Exchange rate regime, crisis, shock adjustment, theory of optimum currency areas, Mundell, Schumpeter, cointegration, competitive depreciations.

JEL Classification: F31, F32
AMS Classification: 62J12, 62P20

1 Introduction

The recent wave of financial, balance of payments and sovereign debt crises has revived the discussion about the appropriate adjustment strategy in the face of asymmetric shocks. In most crisis events such as the 1997/98 Asian crisis, the 1998 Japanese financial crisis, the 1998 Russian flu, the 2001 collapse of the Argentine currency board and even the US subprime market crisis, the crisis countries opted for monetary expansion and depreciation as crisis solution strategies. In contrast during the most recent European sovereign debt crises a set of European crisis countries opted for staying in the Economic and Monetary Union (EMU) (the EMU crisis countries) or maintaining tight exchange rate pegs to the euro (such as the Baltic countries). The consequence was a strong pressure to curtail government expenditure and to cut nominal wages.

The different adjustment strategies in the face of crisis based on inflation or deflation are embedded into different theoretical frameworks. Keynes [13] and Mundell [17] favoured monetary expansion and depreciation to provide a quick fix for missing international competitiveness and high unemployment. In contrast, Schumpeter [21] and Hayek [8] had stressed the positive long-term growth effects of wage and price cuts to increase the marginal efficiency of private investment. In the context of the choice of the exchange rate regime, Mundell [17] stressed the benefits of exchange rate adjustment in the face of asymmetric shocks to stimulate growth (short-term), whereas McKinnon [15] highlighted the role of fixed exchange rates for macroeconomic stabilization and therefore as a tool for preserving the long-term growth performance. From an economic historical perspective Bordo [2] argues that the exchange rate stability of the gold standard has fostered global trade and growth, whereas Eichengreen [6] sees the early exit from the gold standard as the reason for UK's early recovery from the world economic crisis.

Empirical studies on the impact of the exchange rate regime on growth have come to mixed results. For instance, Levy-Yeyati and Sturzenegger [11] who examine the impact of the exchange rate regime on growth for a sample of 183 countries in the post-Bretton-Woods era (1974-2000) based on a pooled regression framework find a negative impact of exchange rate stability on growth for emerging market economies. In contrast, Maurel

[1]CERGE-EI, Charles University, Politických vězňů 7, 111 21 Prague Czech Republic, evzen.kocenda@cerge-ei.cz.
[2]CES, Université Paris 1, Panthéon-Sorbonne, 106-112 Boulevard de l'Hopital,75647 Paris cedex 13, France, mathilde.maurel@univ-paris1.fr.
[3]Gunther Schnabl, University of Leipzig, Grimmaische Straße 12, 04109 Leipzig, Germany, schnabl@wifa.uni-leipzig.de.

and Schnabl [16] use static and dynamic panel estimations to find a positive impact of exchange rate stability on growth for a set of 60mostly emerging market economies.

We aim to augment this literature by isolating the long-term and short-term growth effects of exchange rate stability / flexibility based on a cointegration framework. This research allows us – with the help of a panel of 60 countries in East Asia, Europe, Latin America and the CIS for the period from 1980 to 2011 – to reconcile Mundell's [17] and McKinnon's [15] view on the impact of exchange rate flexibility/volatility in the face of asymmetric shocks.

2 Short-Term and Long-Term Growth Effects of (Non-) Exchange Rate Adjustment

From 2008 to 2012 European debt crisis has revealed the different adjustment strategies to asymmetric shocks and crisis. When many European periphery countries were hit by bursting bubbles, the reversal of capital inflows and (near to) unsustainable debt levels, the membership in the EMU barred the way towards depreciation as a quick fix for the adjustment of unit labour costs to regain international competitiveness.

The loss of an independent monetary policy and the exchange rate as adjustment tools to asymmetric shocks made price and wage adjustments necessary, which amplified the crisis and triggered different policy responses. Whereas Ireland (like the Baltic countries in the east of the European Union) embarked on doughty adjustment measures in the private and public sector, in Greece political resistance retarded reforms. The delayed reforms in Greece were reflected in substantial rescue packages and rising imbalances in the TARGET2 mechanism (Sinn and Wollmershäuser [20], Abad et al. [1]), which both provided a substitute for pre-crisis private capital inflows as financing mechanism of persistent current account deficits.

The real exchange rate adjustment, which is needed to rebalance the current account position of crisis countries, is framed by the theories of optimum currency areas. Mundell [17] assumed that countries with a high probability of asymmetric shocks need to preserve the exchange rate as an adjustment mechanism to stabilize growth, if labour markets are rigid. This reflects the Keynesian assumption of (short-term) price and wage rigidity and the crucial role of the government for macroeconomic stabilization. In contrast, McKinnon [15] argued that in small and open economies a fixed exchange rate needs to serve as a macroeconomic stabilizer by absorbing nominal shocks to promote growth. To maintain a fixed exchange rate, sufficient price and wage flexibility is necessary, which is line with Hayek's [8] and Schumpeter's [21] notion that declining prices and wages are the prerequisite for a robust recovery after crisis.

The growth effects of the crisis adjustment strategies based on exchange rate or wage adjustment has a short-term and a long-term dimension as well as a goods market and a capital market perspective. Keynes [13] stressed the short-term dimension with a focus on goods markets. Depreciation in the case of crisis can help to "jumpstart" the economy as monetary expansion and real depreciation restore instantaneously the international competiveness. Real wages decline without burdensome wage negotiations as inflation increases, in particular via the import price channel in timely manner.

Hawtrey [12], who dedicated his academic work to the deflationary consequences of the return to the gold standard, pioneered the financial market based arguments for a monetary expansion during crisis. Interest rate cuts and liquidity injections into the financial sector help to stabilize the financial system. Low-costs credit prevents a monetary tightening during crisis due to increasing risk perception in the private banking sector. The monetary expansion prevents a credit crunch and the dismantling of investment projects (which lead into a cumulative process downwards). New investment during the crisis is encouraged which helps to speed up the recovery. As a fundamental restructuring process in the economy is prevented, dire wage adjustment becomes dispensable, which helps to maintain economic activity via the consumption channel.

Hayek [8] and Schumpeter [21] help to understand the long-term growth effects of crisis therapy via monetary and depreciation, when restructuring and structural reforms are postponed. In Schumpeter's ([21]: 350) real overinvestment theory the recession is a process of uncertainty and disorder, which forces a reallocation of resources on the enterprise sector ("cleansing effect"). The reallocation is a pre-requisite for long-term growth for the following reasons: Speculative investment is abandoned. Inefficient enterprises leave the market. The efficiency of the remaining enterprises is strengthened. New enterprises, products and production processes emerge at the cost of old ones. An exchange rate depreciation due to a crisis is an impediment to long-term growth as the

"unadapted and unlivable" persists (Schumpter [21]: 367).[4] Hence, an exchange rate regime matters, as a fixed exchange rate (or membership in a monetary union) imposes the need for reforms.

The monetary overinvestment theory of Hayek [8] allows approaching the long-term growth effects of monetary expansion and depreciation from a capital market perspective. The pivotal point is the impact of monetary expansion on the marginal efficiency of investment. During the upswing artificially low interest rates set by the central bank encourage investment with declining marginal efficiency. When rising inflation urges the central bank to lift interest rates again, investment projects with a low marginal return have to be dismantled. The resulting cleansing effect is the prerequisite for a sustained recovery. If, however, the central bank responds to the crisis by decisive interest rate cuts, investment projects with low marginal returns, i.e. a distorted production structure; are conserved (Hoffmann and Schnabl [9], [10]). A structurally declining interest rate deprives the interest rate of its allocation mechanism – which separates high-return investment from low-return investment – thereby putting a drag on long-term growth (Schnabl [19]).

This implies that – depending on the theoretical framework – exchange rate adjustment in face of crisis has a positive or negative growth effect depending on the time horizon. As stressed by Mundell [17] the question of if the exchange rate regime has a positive or negative effect on the growth performance of countries may – with an asymmetric world monetary system – further hinges on the degree of business cycle synchronization with the (potential) anchor country. Because of underdeveloped goods and capital markets small and open economies have an inherent incentive to stabilize exchange rate versus the currency of a large anchor country (usually the dollar or the euro) (Calvo and Reinhart [4]). If business cycles are synchronized with the anchor country, the monetary policy of the anchor country will be in line with the macroeconomic needs of the small open economy. If, however, business cycles are idiosyncratic, there is a larger need to create a growth stimulus via depreciation.

Previous papers have tested for the overall growth effects of exchange rate flexibility, partially contingent on business cycle synchronization. We augment this literature by separating between the long-term and short-term growth effects of exchange rate flexibility based on a sample of 60 small, open emerging market economies based on an error correction framework.

3 Sample, Volatility Measures and Business Cycle Correlation

To trace the short-term and long-term impact of exchange rate flexibility, i.e. exchange rate volatility, on growth, we choose five country groups for which the choice of the appropriate exchange rate regime has been high on the political agenda: In the EU15 as well as in Central, Eastern and Southeastern Europe (Emerging Europe), the discussion about membership in the EMU and/or the optimum degree of exchange rate stability against the euro continues to be high on the political agenda. Empirical results of Kočenda and Poghosyan [14] imply that to contribute to further stability of the domestic currency, the new EU members should strive to implement stabilization policies aimed at achieving nominal as well as real convergence with the core EU members since both real and nominal factors play important roles in explaining the variability of their foreign exchange risk premium.

The discussion about the pro and cons of EMU membership and exchange rate stability against the euro was revived during the most recent crisis. In East Asia and South America the optimum degree of exchange rate stability against the dollar continues to be discussed, in particular since the Asian crisis and drastic US interest rate cuts following the subprime crisis. Most recently, Japan, China and Brazil have been involved in a discussion on "currency wars" and competitive interest rate cuts (McKinnon [15]). In the Commonwealth of Independent States, Russia's move towards a currency basket and the depreciation of the CIS currencies during the recent crisis has revived the question about the optimum exchange rate policy. In this context, the choice of the anchor currency and therefore the degree of business cycle synchronization with the anchor country play an important role.

The five country groups include all countries of the respective region excluding microstates – which may bias the sample towards a very high positive effect of exchange rate stability on growth (Rose [18]) – and countries with insufficient data such as Turkmenistan, Uzbekistan, and Guyana etc. This brings us to a sample size of 60 countries. They are grouped into the regions with prevailing anchor currencies and, hence, the reference countries for measuring business cycle correlation. For the countries in East Asia, South America and the CIS the dollar has been the prevailing target of exchange rate stabilization. Business cycle correlation is measured versus the US. For the European countries before the introduction of the euro in 1999, the German mark has been the dominant anchor currency. Since then, the euro has become the natural anchor for the European non-EMU coun-

[4] Schumpeter's [21] argument which has been designed for the private enterprise sector can be applied for the government sector as well. A strong recession will trigger only structural reforms if there are restrictions on fiscal and monetary expansion in place.

tries. Exchange rate volatility is measured against the German mark before 1999 and against the euro after 1999. Once a country has entered the EMU the proxy for exchange rate volatility is set to zero. Business cycle correlation in Europe is measured versus Germany, which is the largest European economy (and therefore a country with a high degree of business cycle correlation with the euro area). For Germany, France as the second largest European economy is used as a reference country to measure business cycle correlation.

The data source is mainly the IMF International Financial Statistics. Missing or inconsistent data are completed and cross-checked with national statistics, mainly national central banks. For all macroeconomic data we use quarterly frequencies. Quarterly GDP growth rates are calculated as year-over-year quarterly growth rates to filter out seasonal pattern and lower the volatility of the transformed series. Quarterly volatility measures for exchange rates are computed based on monthly data. The sample period starts in January 1994 to avoid putting in the analysis the first years of the nineties, which for most of the European and CIS countries implied high volatility linked to the transition process.

4 Empirical Analysis

Given the different time dimensions of economic theories concerning the impact of exchange rate stability/flexibility on growth, the issue of crisis adjustment via exchange rate flexibility is an empirical one. In contrast to previous papers, we aim to address the question about the optimum exchange rate regime by identifying the impact of exchange rate flexibility on both short-run and long-term growth. Hawtrey [12], Keynes [13] and Mundell [17] argued that exchange rate adjustment during crisis increases the (short-term) growth performance of countries with a high likelihood of asymmetric shocks. In contrast, for Schumpeter [21], Hayek [8] and McKinnon [15] exchange rate stability during crisis enhances the need for internal adjustment, what is seen as a prerequisite for dynamic long-term growth. The foregoing empirical analysis aims to distinguish the long-term and short-term effect of exchange rate flexibility on growth, which may possibly, reconcile both strands of literature by attributing a time dimension to them.

4.1 Model Specification and Estimation Procedure

Equation (1) is our benchmark equation. The explanatory variables are exchange rate volatility, inflation, interest rate, and openness and trend as additional control variables.

$$w_{it} = a + b_1 ERV_{it} + b_2 ERV_{it}^{assy} + c\ inflation + d\ interestrate + controlvariables + ec\big(\beta_1 ERV_{it} + \beta_2 ERV_{it}^{assy} + \gamma\ inflation + \delta\ interestrate\big) \quad (1)$$

In specification (1) w_{it} are the quarterly real growth rates from 1994 to 2010. ERV_{it} stands for the three measures of exchange rate volatility: standard deviations, mean of percent exchange rate changes against the anchor currency, and *z*-score (see Maurel and Schnabl [16] for details). *Interest rate* stands for short-term interest rates of the respective reference country as one of the most important determinants of global growth. Average *inflation*, proxied by the average of monthly year-over-year changes of the consumer price index controls for (negative) growth effects originating in macroeconomic instability, which is linked to exchange rate volatility. We control for the fact that countries with asymmetric business cycles face bigger constraints to achieve the same growth by using the dummy *Dbbc* (see earlier explanation in Section 3). Combined dummy variable ERV_{it}^{assy} is constructed as ERV_{it} times *Dbbc* and is set equal to one for countries characterized by asymmetric business cycles, and zero otherwise. The term after error correction term *ec* in parentheses refers to the short-run adjustment while the first part of the equation displays the long-term relationship. Control variables in the long-run equation are *trade openness*, *time*, and *squared time*, to better fit the non-linear pattern of growth especially in transition countries during the first years of the nineties. We assume that those variables make sense for assessing the long-run economic performance while they do not matter in the short-run.

There is a large number of other potential explanatory variables like investment, government spending, schooling etc. which could increase the fit of the model, but also generate endogeneity (for instance investment and growth) and multicollinetarity bias (for instance between government spending and inflation) (De Grauwe and Schnabl [5]). Therefore, we opt for a parsimonious specification, restricted to the control variables mentioned above.

Our objective is threefold: to highlight the impact of exchange rate volatility on growth, to quantify the weight of countries with asymmetric business cycles (which we call asymmetric countries) in this impact, and to disentangle the short-run versus long-run effect of exchange rate volatility. Our prior is that the impact of exchange rate volatility should be positive in the short-run, especially for asymmetric countries, but negative in the long-run.

An error-correction version of specification (1) is estimated and allows concluding that the effect of exchange rate volatility if negative in the long-run, while it is positive in the short-run, especially for countries with a high probability of asymmetric shocks, as emphasized by Mundell [17]. The results are obtained via the pooled mean group estimation. The restriction of the pooled mean group against the mean group model is tested by performing Hausman tests.

4.2 Estimation Results

We present results of our estimation in Table 1. The null hypothesis of equality of coefficients cannot be rejected at 1% level. This evidence favors the pooled mean group model against the mean group estimator. The pooled mean group results are consistent with those reported in other studies, they hold whatever indicator for exchange rate volatility (standard deviation, average yearly change, and z-score) is used. Their effect is economically important. They provide evidence of a significant negative long-run correlation between exchange rate volatility and growth as well as a positive short-run correlation between exchange rate volatility and growth. They are in line with Furceri and Zdzienicka [7] who show that flexible exchange rates attenuate crises in the short and medium term but tend to have a negative effect on long-term growth, given that the cleansing effects did not materialize. All long-run coefficient estimates are highly significant (at 1 and 5% significance levels) and display the expected signs. A sound macroeconomic policy keeping inflation low is conducive to growth, while lower interest rates imply more opportunities for investment, which translates into higher growth.

Table 1 Short- and Long-Run Effects (pooled mean group estimators)

	Exchange Rate Volatility Measures		
Variable	Average change	Standard deviation	z-score
Long-run coefficients			
Time	-0.0019*** (0.001)	-0.0024*** (0.001)	-0.0023*** (0.001)
$(Time)^2$	0.0000* (0.000)	0.0000** (0.000)	0.0000** (0.000)
Trade openness	0.0331*** (0 .008)	0.0497*** (0.013)	0.0530*** (0.013)
ER volatility	-0.1825 (0.121)	-0.5971*** (0.122)	-0.4916*** (0.096)
ERvolatility (asymmetric)	-0.5382*** (0.204)	0.1497 (0.187)	0.0921 (0.144)
Inflation	0.0006 (0.019)	-0.0246 (0.021)	-0.0033 (0.020)
Interest rate	-1.3052*** (0.165)	-1.2316*** (0.156)	-1.2779*** (0.158)
Short-run coefficients			
Error correction coefficient	-0.2644*** (0.018)	-0.2618*** (0.014)	-0.2640*** (0.014)
Trade openness	0.0830 (0.078)	-0.0760 (0.055)	-0.0803* (0.042)
ERvolatility	0.0289 (0.109)	0.1534** (0.067)	0.1256** (0.057)
Inflation	0.0396* (0.021)	0.0322* (0.019)	0.0284 (0.019)
Interest rate	1.6898*** (0.179)	1.6823*** (0.190)	1.6862*** (0.181)
Intercept	0.0346*** (0.003)	0.0360*** (0.003)	0.0358*** (0.003)

Significance levels at 1, 5, and 10% are denoted by ***, **, and *, respectively.
Note: ER volatility denotes one of the three exchange rate volatility measures used; e.g. average yearly change, standard deviation, and z-score. Numbers in parentheses are standard errors.

With respect to the short-term elasticities, the pooled mean group results suggest that interest rates and average inflation have highly significant short-run effects on growth. Exchange rate volatility is negatively signed, but its magnitude is significantly lower than for the long-run. The interaction terms capture the impact of exchange rate volatility on growth for asymmetric countries. They are significant, meaning that the impact of exchange volatility is magnified for asymmetric countries. Those findings support the view of exchange rate flexibility helping to smooth out asymmetric shocks only in the short run.[5] They do not serve the objective of a dy-

[5]Bubák et al. [3] document the existence of volatility spillovers between the Central European foreign exchange markets as well as the fact that exchange rate volatility increases in medium-term for those countries with troubled financial sector development.

namic long-term growth performance. The estimated adjustment parameter is negatively signed and significant, implying that short-run deviations from the long-run growth path dampen over time.

5 Conclusion

With the European sovereign debt crisis a controversial discussion concerning the optimum monetary policy and exchange rate strategy to asymmetric shocks has reemerged. We have aimed to derive from our econometrical exercise for a panel of 60 countries a policy recommendation for crisis countries. Our estimation results have provided evidence that exchange rate adjustment stimulates growth in the short-term, but puts a drag on the long-term growth performance. As the overall effect is negative, the policy implication is to keep exchange rates stable to promote growth via price and wage flexibility, in the spirit of Schumpeter [21], Hayek [8] and McKinnon [15].

Based on our findings we recommend to the crisis countries consequent structural reforms and real wage cuts. Painful restructuring and declining output today are likely to be rewarded with a robust economic recovery and rising income in the future. In contrast, monetary expansion and depreciation as a crisis solution strategy can be expected to provide short-term relief, but long-term pain.

Acknowledgements

We benefited from comments by presentation participants at Prague and Shanghai. Evžen Kočenda acknowledges research support from the GAČR grant No. 403/11/0020. The usual disclaimer applies.

References

[1] Abad, J., Löffler, A., Schnabl, G., and Zemanek, H.: Fiscal Divergence, Current Account and TARGET2 Imbalances in the EMU. *University of Leipzig Wirtschaftswissenschaftliche Fakultät Working Paper* **105** (2012).
[2] Bordo, M.: Globalization in Historical Perspective. *Business Economics* **1** (2002), 20–29.
[3] Bubák, V., Kočenda, E., and Žikeš, F.: Volatility Transmission in Emerging European Foreign Exchange Markets. *Journal of Banking and Finance* **35** (2011), 2829–2841.
[4] Calvo, G., and Reinhart, C.: Fear of Floating. *The Quarterly Journal of Economics* **117** (2002), 379–408.
[5] De Grauwe, P., and Schnabl, G.: Exchange Rate Stability, Inflation and Growth in (South) Eastern and Central Europe. *Review of Development Economics* **12** (2008), 530–549.
[6] Eichengreen, B.: *Financial Crises and What to Do About Them*. Oxford, 2002.
[7] Furceri D., and Zdzienicka A.: How Costly Are Debt Crises? *IMF Working Papers* **11/280** (2011).
[8] Hayek, F. Von: *Monetary Nationalism and International Stability*. London, 1937.
[9] Hoffmann, A., and Schnabl, G.: A Vicious Cycle of Financial Market Exuberance, Panics and Asymmetric Policy Response – An Overinvestment View. *The World Economy* **34** (2011), 382–403.
[10] Hoffmann, A., and Schnabl, G.: National Monetary Policy, International Economic Instability and Feedback Effects – An Overinvestment View. *Global Financial Markets Working Paper* **19** (2011).
[11] Levy-Yeyati, E., and Sturzenegger, F.: To Float or to Fix: Evidence on the Impact of Exchange Rate Regimes on Growth. *American Economic Review* **12** (2002), 1–49.
[12] Hawtrey, R.: The Gold Standard. *Economic Journal* **3** (1919), 428–442.
[13] Keynes, J. M.: *The General Theory of Employment, Interest, and Money*. London, 1936.
[14] Kočenda, E., and Poghosyan, T.: Macroeconomic Sources of Foreign Exchange Risk in New EU Members. *Journal of Banking and Finance* **33** (2009), 2164–2173.
[15] McKinnon, R. I.: Optimum Currency Areas. *American Economic Review* **53** (1963), 717–724.
[16] Maurel, M., and Schnabl, G.: Keynesian and Austrian Perspectives on Crisis, Shock Adjustment, Exchange Rate Regime and (Long-Term) Growth. *Forthcoming in Open Economies Review*.
[17] Mundell, R.: A Theory of Optimum Currency Areas. *American Economic Review* **51** (1961), 657–665.
[18] Rose C.: Finance Constraints and Inventory Investment: Empirical Tests with Panel Data. *Macroeconomics EconWPA* **38** (2004).
[19] Schnabl, G.: Exchange Rate Volatility and Growth in Emerging Europe and East Asia. *Open Economies Review* **20** (2009), 565–587.
[20] Sinn, H-W., and Wollmershäuser, T.: Target Loans, Current Account Balances and Capital Flows: The ECB's Rescue Facility. *NBER Working Paper* **17626** (2011).
[21] Schumpeter, J.: *Theorie der Wirtschaftlichen Entwicklung*. Berlin, 1911.

A modification of Kaldor-Kalecki model and its analysis

Jan Kodera[1], Jarmila Radová[2], Tran Van Quang[1]

Abstract. This paper studies the investment cycle as an endogenous phenomenon using a modification of so called Kaldor-Kalecki's model as a tool of research. Our aim is to show that the (Kaldor-Kalecki) model which contains such features as non-linearity in the investment function, a delay between an investment decision and its delivery, has an ability to produce more complex behaviour of its variables, i.e. periodical or non-periodical oscillations. As the model is a system of two differential equations with delay which the common solvers are not able to reliably deal with so far, a special solver is developed by authors for this case and will be presented in the article.

Keywords: Kaldor-Kalecki model, system of non-linear differential equations, investment cycle, limit cycle, complex dynamics

JEL classification: C44
AMS classification: 90C15

1 Introduction

Economic cycles in traditional macroeconomics are closely connected with investment fluctuations. Examining thoroughly macroeconomic theories of endogenous cycle, we can find that two possible explanations of this phenomenon. It is the delay between an investment decision and the delivery of investment and the non-linearity in the investment function. The former approach can be found in the original paper from Kalecki [9], which later was also presented in Allen's work [1]. In his model, Kalecki assumes that there is a gestation period which is the time between the moment when an investment decision is made and the time of delivering of the finished real investment. The process of construction requires time and its mathematical description leads to a differential equation with delay which may exhibit more complex dynamics of model variables. In Allen's book, both the elder version of Kalecki's model (1935) and its later version proposed in (1943) and (1954) can be found and they are a generalisation of the original version. Both Kalecki's models describe the capital dynamics with the help of a linear differential equation with delay capable of generating periodical oscillation of capital.

The approach based on non-linearity of investment function usually uses a system of two or more differential equations. As we know, non-linear deterministic dynamic systems could display non-linear oscillations. The theory of non-linear deterministic systems can be found in Guckenheimers and Holmess famous work [5]. Their work attracted many followers, for example: Kuznetsov [12], Perko [14]. From the point of view of non-linear system theory, an economy is a non-linear system. This interpretation of economic systems is fundamental because according to it, the origin of economic fluctuations results from the inside structure of the system, not as a consequence of irregular external shocks as the real business cycles theory suggests.

Another approach to explain investment cycles can be found in Kaldor's original model in his seminal work [8]. While Kalecki's model is reduced to one differential equation with delay describing the capital formation, Kaldor's original idea is to study the evolution of production and capital formation. Kaldor suggests that the treatment of savings and investment as linear curves simply does not correspond to empirical reality. He assumed that investment and savings are both positive non-linear and non-convex functions of output (income) and that investment depends negatively on capital while savings dependence on capital is positive. From the non-convex shape of both curves and from their movement depending on capital he derived endogenous cyclical behaviour of production and capital.

Kaldor original approach has many modifications and improvements. Chang and Smyth [3] were the first

[1]University of Economics, Prague, Department of Statistics and Probability, Nám. Winstona Churchilla 4, Praha 3
[2]University of Economics, Prague, Department of Banking and Insurance, Nám. Winstona Churchilla 4, Praha 3,
email: {kodera,radova,tran}@vse.cz

authors who translated the original Kaldor's idea into a rigorous form. Modern forms of Kaldor's model and a detailed analysis of its solution could be found in works of Lorenz [13] or Gabisch and Lorenz [4]. Later, Kodera and Vosvrda [10] included Kaldor's model into a model of price dynamics and developed a model in which the dynamics of price level is strongly affected by oscillations of production and capital in the real sector.

To develop further the essential idea of endogenous business cycle theory, in our contribution we will connect both approaches (Kalecki and Kaldor) and modify the Kaldor-Kalecki model as a system of two differential equations with delay. Though this idea is not new thing and a high amount of papers devoting to this problem, for example, it can be found in Kaddar and Talibi Alaoui [7] or Krawiec and Szydlowski [11], our contribution will take a step further. Although we use the same general formulation of Kalecki-Kaldor's model as in the above mentioned papers, unlike these authors, we introduce a specific investment function of logistic form as one source of linearity. Moreover, without losing the general validity of results and conclusions, the savings function in our model is a linear function of logarithm of production. Above all, unlike other authors in mentioned papers we pay more attention to solving the Kalecki-Kaldor's model. Despite the fact that there are some toolboxes solving differential equations with delay, we do not find them as appropriate and reliable for this purpose, and therefore, we develop our own software tool for solving Kalecki-Kaldor's model as a system of delay differential equations.

The rest of our paper is structured as follows. In the next section, we are going to present the original Kaldor's model and show the cyclical behavior of solution for production and capital on a numerical example. In the third section the formulation of Kalecki-Kaldor's model is given. In fourth section, we will explain the method for numerical solution of Kalecki-Kaldor's model and apply it on a calibrated model. In the final section a short discussion as well as some features of obtained solution will be given.

2 Kaldor's Model

In this section, we will give a brief description of Kaldor's model with non-linear investment function. The model consists of two differential equations

$$\dot{Y}(t) = \alpha[I(Y(t),K(t)) - S(Y(t))], \tag{1}$$
$$\dot{K}(t) = I(Y(t),K(t)) - \delta K(t), \tag{2}$$

where $Y(t)$ denotes real production, $K(t)$ denotes capital, $I(Y(t),K(t))$ labels investment which is an increasing function of production and is decreasing in capital. The savings function is an increasing function of production and is denoted as $S(Y(t))$. The second equation describes capital formation which is given by the difference between investment I and capital consumption $\delta K(t)$ where δ denotes the rate of deprecation.

The investment function will be presented in a specific form. Let's leave aside for a moment that it is a function of time and assume it is a product of two functions $J(Y,K)$ and Y. Then $J(Y,K)$ is in fact an investment-product ratio frequently called the propensity to invest. It is common to assume that it is an increasing function of productivity of capital, therefore we have $J(Y,K) = J(Y/K)$. In our model, for convenience, we use the logarithmic form and the original function $J(Y,K)$ can be rewritten as $J(Y/K) = J(\mathrm{e}^{y-k}) = i(y-k)$, where $y = \log Y$, $k = \log K$. As the logistic function is chosen as the functional form for the propensity to invest $i(t)$, it can be defined as follows

$$i(y(t)-k(t)) = \frac{ai_0}{bi_0 + (a-bi_0)\mathrm{e}^{-a(y(t)-k(t))}}. \tag{3}$$

Though the choice of the propensity to invest function seems to be rather arbitrary and speculative, it meets all requirements of a function describing the propensity to invest. The graph of this function for parameters $a = 2$, $b = 8$, $i_0 = 1/8$ is shown on the Fig. 1. Further, it also has maximum lower bound and minimum upper bound (0 and a/b respectively. The investment function then is of the following form:

$$I(Y(t),K(t)) = \frac{ai_0}{bi_0 + (a-bi_0)\mathrm{e}^{-a(y(t)-k(t))}}Y(t). \tag{4}$$

By the same token, the real savings function is defined as a product of the propensity to save s and the production. We assume that the propensity to save is increasing in logarithm of production. This assumption shows that the propensity does not grow as quickly as production, but much slower. So the mathematical expression for the propensity to save function is

$$s(\log Y(t)) = s_0 + s_1 \log Y(t) = s_0 + s_1 y(t). \tag{5}$$

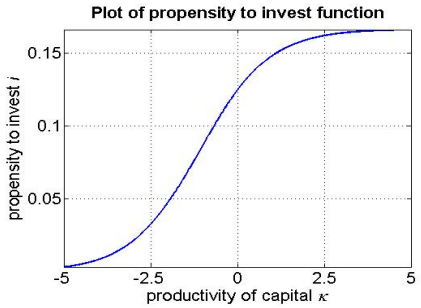


Figure 1 Logistic propensity to invest function

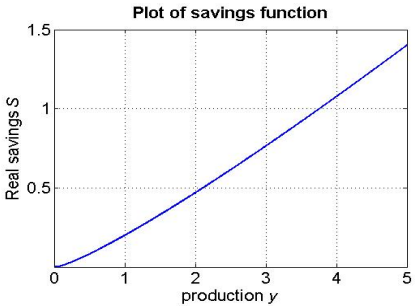


Figure 2 Real savings function

The savings function then is:

$$S(y(t)) = (s_0 + s_1 y(t))Y(t). \tag{6}$$

The real savings function with $s_0 = 0.2$ and $s_1 = 0.05$ is displayed in Fig. 2.

Using assumptions being made about investment and savings functions in (1) and (2), we get

$$\dot{Y}(t) = \alpha[i(y(t) - k(t))Y(t) - s(y(t))Y(t)], \tag{7}$$
$$\dot{K}(t) = i(y(t) - k(t))Y(t) - \delta K(t). \tag{8}$$

Dividing equation (7) by Y and (8) by K, we have

$$\dot{y}(t) = \alpha[i(y(t) - k(t)) - s(y(t))], \tag{9}$$
$$\dot{k}(t) = i(y(t) - k(t))\mathrm{e}^{y(t)-k(t)} - \delta, \tag{10}$$

where

$$\dot{y}(t) = \frac{\dot{Y}(t)}{Y(t)}, \qquad \dot{k}(t) = \frac{\dot{K}(t)}{K(t)}.$$

Substituting (3) and (5) into (9) and (10), we have

$$\dot{y}(t) = \alpha\left[\frac{ai_0}{bi_0 + (a - bi_0)\mathrm{e}^{-a(y(t)-k(t))}} - (s_0 + s_1 y(t))\right], \tag{11}$$
$$\dot{k}(t) = \frac{ai_0}{bi_0 + (a - bi_0)\mathrm{e}^{-a(y(t)-k(t))}} e^{y(t)-k(t)} - \delta. \tag{12}$$

The plot of the solutions of these two equations for $a = 2$, $b = 8$, $i_0 = 1/8$, $s_0 = 0.2$, $s_1 = 0.05$ and $\delta = 0.1$ is displayed on Fig. 3 and the phase portrait is shown on Fig. 4.

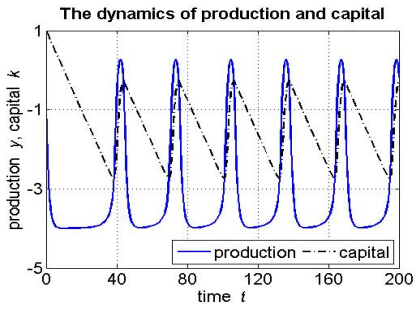


Figure 3 The evolution of production and capital

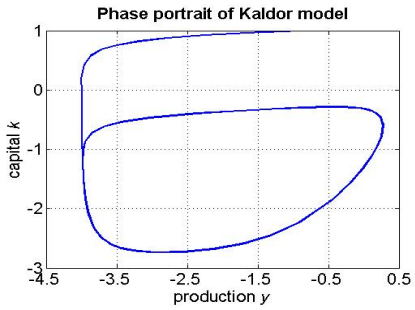


Figure 4 Phase portrait of Kaldor model

3 Kaldor-Kalecki model

Kalecki [9] in his famous article generalised Tinbergen's [15] idea of industrial investment cycles based on the ship-building model. Tinbergen observed gestation period from decision to invest to putting the investment to operation. He used a delay-differential equation for the description of this problem. Kalecki elaborated Tinbergen's original

model using the idea of accelerator-multiplier joint activity. Accelerator-multiplier principle and the delay between investment decision and its delivery is expressed in delay-differential equation which models the process of capital formation. Tinbergen and Kalecki's idea of gestation period was utilized in modification of original Kaldor's model and resulted in Kaldor-Kalecki's model.

Kaldor-Kalecki's model takes into account a delay between investment decision and the delivery of investment. Capital formation denoted as $\dot{K}(t)$ is a difference between the investment decision performed in time $t-\theta$, $\theta > 0$ and capital consumption. Parameter θ stands for a gestation period. The first equation in Kaldor-Kalecki's model is the same as the one in original Kaldor's model. The decision to invest or the demand for investment is constituted in time t. The second equation is different because the investments delivered in time t are the result of investment decisions taken at time $t-\theta$. The Kaldor's system of differential equations (1) and (2) is modified to a new system as follows

$$\begin{aligned} \dot{Y}(t) &= \alpha[I(Y(t),K(t)) - S(Y(t))], \\ \dot{K}(t) &= I(Y(t-\theta),K(t-\theta)) - \delta K(t). \end{aligned}$$

As the assumptions about investment and savings functions remain unchanged the same, using

$$I(Y(t),K(t)) = i(y(t)-k(t))Y(t), \quad S(Y(t)) = s(y(t))Y(t),$$

we get the following system of differential equations

$$\begin{aligned} \dot{Y}(t) &= \alpha[i(y(t)-k(t))Y(t) - s(y(t))Y(t)], \\ \dot{K}(t) &= i(y(t-\theta)-k(t-\theta))Y(t-\theta) - \delta K(t). \end{aligned}$$

Dividing the first equation by $Y(t)$ and the second one by $K(t)$, we have

$$\begin{aligned} \dot{y}(t) &= \alpha[i(y(t)-k(t)) - s(y(t))], \\ \dot{k}(t) &= i(y(t-\theta)-k(t-\theta))\mathrm{e}^{y(t-\theta)-k(t)} - \delta. \end{aligned}$$

Using (3) and (5)and substituting them into the two differential equations above, we get

$$\dot{y}(t) = \alpha\left[\frac{ai_0}{bi_0 + (a-bi_0)\mathrm{e}^{-a(y(t)-k(t))}} - (s_0 + s_1 y(t))\right], \tag{13}$$

$$\dot{k}(t) = \frac{ai_0}{bi_0 + (a-bi_0)\mathrm{e}^{-a(y(t-\theta)-k(t-\theta))}} e^{y(t-\theta)-k(t)} - \delta. \tag{14}$$

Kaldor-Kalecki's model is a typical system of two delay differential equations. We will use the same set of parameters as in the case of original Kaldor's model to calibrate it.

4 Solving the model

Kaldor Kalecki's model as derived above is a system a two delay differential equations and in our paper we will solve it for a set of well chosen parameters. Unlike systems of ordinary differential equations which can be numerically solved relatively easily using the Runge-Kutta method, the presence of the lags in the right hand side makes systems of delay differential equations much more difficult to deal with. The difficulty of solving a delay differential equation is shown in the following example. Let's solve this simple DDE equation:

$$\dot{y}(t) = y(t-1). \tag{15}$$

Without the delay, this equation is an ODE equation of first order and can be solved easily both analytically and numerically [1]. We just need one initial condition to exactly identify the solution from a set of solutions. The first deviation from the ODE equation is that in order to identify the solution of (15), in stead of one initial condition, we need to know a whole series of initial conditions up to τ which is called history. Suppose that for $-1 \le t \le 0$ $y(t) = 1$. Then equation (15) becomes $\dot{y}(t) = 1$ with the initial condition $y(0) = 1$ and therefore for $0 \le t \le 1$, then the solution is $y = t + 1$. For $1 \le t \le 2$, then equation (15) becomes $\dot{y}(t) = t - 1 + 1 = t$ with the initial condition $y(1) = 2$ and the solution of 15 is $y = \frac{1}{2}t^2 + \frac{3}{2}$ and so on. Solving a delay differential equation results in solving a series of ODE equations within bounded intervals. As a result we get a solution with an important feature. The solution of equation (15) is continuous, but at $t = 0$ it has $\dot{y}(0_-) \neq \dot{y}(0_+)$, at $t = 1$ $\ddot{y}(1_-) \neq \ddot{y}(1_+)$ and generally at $t = k$ it has $y^{(k+1)}(k_-) \neq y^{(k+1)}(k_+)$. These points are discontinuous points and the solution

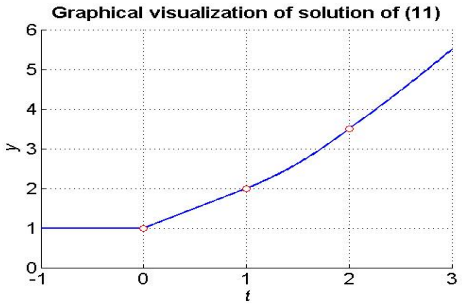


Figure 5 Solution of DDE equation (15)

of equation (15) is shown in Fig. 5. Since a DDE equation can be considered as a sequence of ODE equations, the Runge Kutta method can be used to numerically solve it. For an ODE equation of form $\dot{y}(t) = f(t, y(t))$, its numerical solution according to Runge and Kutta is as the following:

$$y_{n+1} = y_n + h \sum_{i=1}^{s} b_i k_i, \tag{16}$$

where h is the step length, b_i is the weight of k_i, s is the corresponding order of RK method and

$$k_1 = hf(t_n, y_n) \tag{17}$$
$$k_i = hf(t_n + c_i h, y_n + a_i k_{i-1}) \quad \text{for } i = 2, ..., s, \tag{18}$$

where c_i and a_i are pre-set coefficients depending on the order s of the method.
In the case of a DDE equation of form $\dot{y}(t) = f(t, y(t), y(t-\tau))$ with history $y(t) = s(t)$ for $t \leq 0$, the Runge - Kutta method is modified to get the solution in the following way:

$$y_{n+1} = y_n + h \sum_{i=1}^{q} b_i k_i, \tag{19}$$

where

$$k_1 = hf(t_n, y_n(t), y(t-\tau)) \tag{20}$$
$$k_i = hf(t_n + c_i h, y_n + a_i k_{i-1}, y(t_n + c_i h - \tau)) \quad \text{for } i = 2, ..., s. \tag{21}$$

Since for $t \geq 0$ the values of $y(t_n + c_i h - \tau)$ where $t_n + c_i h - \tau$ is the delay argument are unknown, we have to calculate them by interpolation. Due to the existence of discontinuities, when interpolating, the support points should be chosen in such way that they do not include discontinuities and the delay argument should be in the center of interval formed by those chosen support points. Numerical methods for solving ODE and DDE equations can be found in [2] and [6].
Using the method described above, Kaldor-Kalecki's system with the same values of parameters used for numerical calibration as in the case of original Kaldor model is numerically solved in the environment of MATLAB. The values of these parameters are $a = 2$, $b = 8$, $i_0 = 1/8$, $s_0 = 0.2$, $s_1 = 0.05$ and $\delta = 0.1$. The result is illustrated in Fig. 6 and Fig. 7. Solving this system, we realize that the dynamics of production and capital is very sensitive to the values of parameters as well as the delay and the system therefore can create a rich set of dynamics.

5 Conclusion

In our paper we have tried to revitalize the traditional Kaldor's and Kalecki's models, combine them together. Further, we introduce nonlinearity into the model in the form of logistic function. As such the model becomes a system of two delay differential equations with a constant delay. We also choose a set of parameters for the model and numerically solve it with Matlab. The solution of the system exhibits very rich dynamics with some chaotic feature. We consider the extension of our current work to two directions very interesting. First, relaxing assumption of one constant delay would make the model more realistic. Second, analyzing the data generated by the model and reconstructing the possible dynamics from real data would be an important step for verifying the validity of the model.

[1] Analytical solution of equation $\dot{y}(t) = y(t)$ is $y = \mathrm{e}^t$ for the initial condition $y(0) = 1$

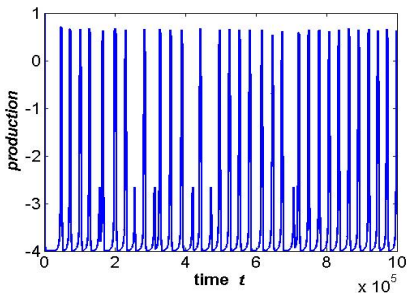


Figure 6 The dynamics of production

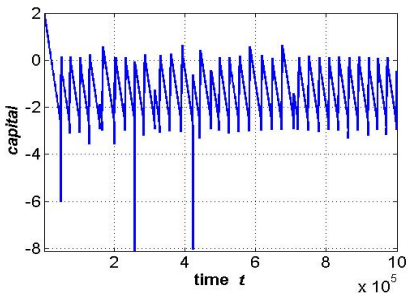


Figure 7 The dynamics of capital

Acknowledgements

The authors are grateful for the financial support of Czech Science Foundation No. P402/12/G097 .

References

[1] Allen, R. D. G.: *Mathematical Economics*. Macmillan, London, 1956.

[2] Bellen, A., Zennaro, M.: *Numerical Methods for Delay Differential Equations*. Oxford Science, Clarendon Press, 2003.

[3] Chang, W. W., Smyth, D. J.: The Existence and Persistence of Cycles in a Non-linear Model: Kaldor's 1940 Model Re-examined. *Review of Economic Studies* **38** (1971), 37–44.

[4] Gabisch, G., Lorenz, H. W.: *Business Cycle Theory: A survey of methods and concepts*. Springer-Verlag, Berlin, 1989.

[5] Guckenheimer, J., Holmes, P.: *Nonlinear Oscillations, Dynamical Systems and Bifurcation of Vector Fields*. Springer-Verlag, Berlin, 1983.

[6] Ismail, F., Al-Khassawneh, R. A.: Solving delay differential equations using embedded singly diagonally implicit Runge Kutta methods. *Acta Mathematica Vietnamica* **33** (2), 95 - 105.

[7] Kaddar, A., Talibi Alaoui, H.: Global Existence of Periodic Solutions in a Delayed Kaldor-Kalecki Model, Nonlinear Analysis. *Modelling and Control* **14** (2009), 463-472.

[8] Kaldor, N.: A Model of the Trade Cycle. *Economic Journal* **50** (1940), 78–92.

[9] Kalecki M.: A Macro-Economic Theory of Business Cycles, *Econometrica* **3** (1935), 327–344.

[10] Kodera, J., Vošvrda, M.: Production Capital Stock, and Price Level Dynamic in the Light of Kaldorian Model. *Acta Oeconomica Pragensia* **15** (2007), 79-87.

[11] Krawiec, A., Szydlowski, M.: The Kaldor Kalecki business cycle model. *Annals of Operations Research* **89** (1999), 89-100.

[12] Kuznetsov, Y. A.: *Elements of Applied Bifurcation Theory*. Springer-Verlag, Berlin, 1998.

[13] Lorenz, H. W.: *Non-Linear Dynamic Economics and Chaotic Motion*. Springer-Verlag, Berlin, 1989.

[14] Perko, L.: *Differential Equations and Dynamical Systems*. Springer-Verlag, Berlin, 2001.

[15] Tinbergen, J.: Ein Schiffsbauzyklus? Weltwirtschaftliches Archiv. Reprinted in Klaasen L.H. et al. *Selected Papers*. Amsterdam, 1959.

Exact approach to the tariff zones design problem in public transport

Michal Kohání[1]

Abstract. An integrated transport system is the way how to provide transport service in the region by integrating all modes of transport. The design of this system is connected with solving of several optimization problems such as coordination of connections in transport modes, optimization of connection supply, minimization of time losses related to the changing of travel connection and the design of the tariff configuration. When designing a tariff system, there are several approaches of designing the tariff. One of them is dividing the region into the tariff zones. The price of traveling in such system is determined by the number of traveled tariff zones by the passenger. In this article we analyze the problem, introduce a mathematical model of the tariff zones design problem based on counting zones and introduce two approaches for determining the quality of solution. We will focus on solving this problem using a universal optimization tool Xpress, on test data set of the selected region. We will compare both approaches in terms of computational time and solution quality.

Keywords: tariff planning, tariff zones design, IP solver, location problem

JEL Classification: C44
AMS Classification: 90C08

1 Introduction

An integrated transport system is the way how to provide transport service in the region by integrating all modes of transport. Also all transport operators in selected region are integrated in this system. The goal of integrated transport system in region is improving the number of transported persons, improve the proportion between individual and public transport in favor of public transport, better coordination between all modes of transport and creating of tariff system, which enables to use one ticket for all modes of transport in the region [3].

The design of this system is connected with solving of several optimization problems such as coordination of connections in transport modes, optimization of connection supply, minimization of time losses related to the changing of travel connection and the design of the tariff configuration.

When designing a tariff system, there are several approaches of designing the tariff. One of them is dividing the region into the tariff zones. The price of traveling in such system is determined by the number of traveled tariff zones by the passenger [4].

In this article we analyze the problem of the tariff system design. In the third chapter we introduce a mathematical model of the tariff zones design problem based on counting zones and introduce two approaches for determining the quality of solution. We focus on solving this problem using a universal optimization tool XPRESS, on test data set of the selected region. We will compare both approaches in terms of computational time and solution quality.

2 Tariff zones design problem

When people use public transport, they usually have to pay for travelling. There are several possibilities how to design ticket prices in public transportation. As was mentioned in [4], the basic and frequently used way is a *distance tariff* system, where the price for a trip depends on the length of the trip. This system is mostly considered as *fair*. If we want to calculate the price for the trip, we need to have the distance between each pair of stations.

Another possibility is the *unit tariff*. It is the simplest tariff system, because in this case all trips cost the same price and are independent on their length. The unit tariff is frequently used in city public transport, but it is not very suitable for regional public transportation, especially for large regions. In this system a short trip between two neighbouring stations leads to the same ticket price as a long trip through the whole system [4].

[1] University of Zilina, Faculty of Management Science and Informatics, Department of Transportation Networks, Univerzitna 1, 01026, Zilina, Slovakia, Michal.Kohani@fri.uniza.sk

Between the unit tariff and the distance tariff there is a *zone tariff* system. When we want to establish a zone tariff, the whole area has to be divided into smaller sub-regions (*the tariff zones*). The price for a trip in a zone tariff system depends only on the starting and the ending zone of the trip. In the zone tariff system there are two possibilities of tariff [4].

If the price is given arbitrarily for each pair of zones, we call the tariff system *a zone tariff with arbitrary prices*. An example for this tariff system is, for instance, the Zilina Regional Integrated Transport System in Slovak Republic, see the tariff prices matrix in the Table 1, or IREDO system in Pardubice and Hradec Kralove region in Czech Republic, see Figure 1.

Prices of basic fare single tickets in ŽRIDS (In Euro)

	To zone 1	**To zone 2**	**To zone 3**	**To zone 4**	**To zone 5**	**To zone 6**	**To zone 7**
From zone 1	x	x	x	1.00	1.00	1.30	1.50
From zone 2	x	x	x	0.90	0.90	1.20	1.40
From zone 3	x	x	x	0.45	0.45	0.75	0.95
From zone 4	1.00	0.90	0.45	0.45	0.45	0.70	0.90
From zone 5	1.00	0.90	0.45	0.45	0.45	0.50	0.70
From zone 6	1.30	1.20	0.75	0.70	0.50	0.45	0.45
From zone 7	1.50	1.40	0.95	0.90	0.70	0.45	0.45

Table 1 Price matrix in Zilina Regional Integrated Transport System *(www.dpmz.sk)*

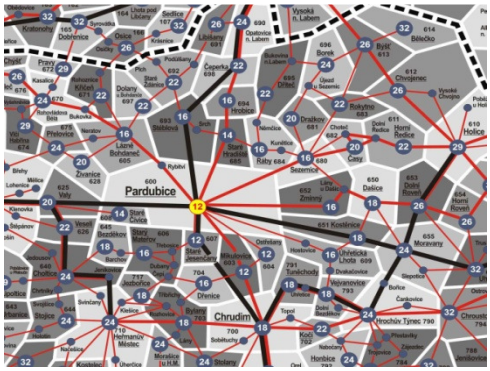


Figure 1 Prices in IREDO system *(www.iredo.cz)*

The second variant of a zone tariff system, popular especially in Germany and Switzerland, is *the counting zone tariff system*. The price of trip in this system is calculated according to the number of crossed zones on the trip. The prices in this system are dependent on the starting and the ending zone of the trip, but trips passing the same number of zones must have the same price. The example of a counting zone tariff system in Zurich region in Switzerland is in Figure 2.

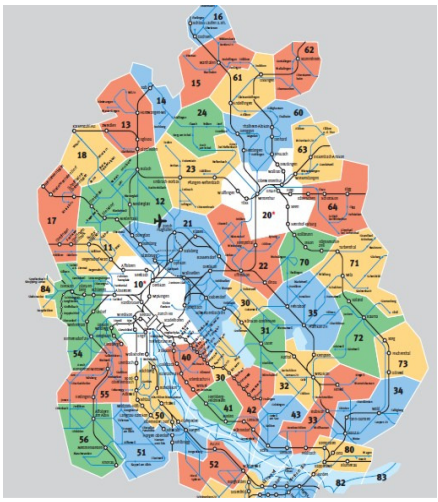


Figure 2 Zone tariff system in Zurich region in Switzerland *(www.zvv.ch)*

When a public transportation company or the regional transportation office wants to change its tariff system to a zone tariff, it has to design the zones and to fix the new fares. The resulting tariff system must be accepted by the customers and does not decrease the income of the company. The goal often is to design the zones in that the new and the old price for most of the trips are as close as possible. This means that neither the public transportation company nor the customers will have major disadvantages when changing the current tariff system to a zone tariff.

Another goal can be to design fair zones. In this case we do not consider the deviation to some old prices, but the deviation from a reference price, for instance one which is considered to be fair, like the distance tariff. In this approach, the public transportation company needs to estimate its new income.

In the available literature we found only the heuristic methods to solve this type of problem [3], [4]. We will formulate the mathematical model of the problem and solve this problem exactly using the IP-solver.

3 Mathematical formulation of the problem

Let all stations in the network of public transport constitute the set I. The station i and j from set I are connected by the edge $(i,j) \in V$, if there is direct connection by public transport line between these two stations. Symbol V 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 stops. 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, how many zones are crossed on this trip. The calculation of the number of crossed zones can be easily replaced by the calculation of crossed zone borders. We assume that the node can be assigned only to one zone and then the border between zones is on the node. We will introduce the binary variable w_{rs} for each existing edge $(r, s) \in V$, 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 traveling between stations i and j. We introduce a_{ij}^{rs}, where the used paths will be observed. a_{ij}^{rs} is equal to 1 if the edge (r,s) will be used for travelling between i and j and 0 otherwise. This calculation will be done before solving the model.

The current or fair price between stations i and j is denoted by c_{ij}. Parameter f represents the price per travelling one zone. New price determined by the number of crossed zones will be calculated as follows (1):

$$n_{ij} = f.\left(1 + \sum_{(r,s)\in V} a_{ij}^{rs} w_{rs}\right) \tag{1}$$

In the creation of the mathematical model of this problem, we were inspired by the model of p-median problems. As we mentioned above, we assume that the node can be assigned exactly to one zone. Then we can 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 variables 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.

According to [4] we introduce two different objective functions. First one will be the maximal deviation between the current or fair price and new price determined by the number of crossed zones for all passengers between i and j. Second one will be the average deviation between current and new price for all passengers.

The mathematical model for criterion maximal deviation can be written in the form:

$$Minimize\ dev_{\max} = \max\{b_{ij}\left|c_{ij} - n_{ij}\right|, i, j \in I\} \tag{2}$$

$$subject\ to \sum_{i\in I} z_{ij} = 1, for\ j \in I \tag{3}$$

$$z_{ij} \le y_i, for\ i, j \in I \tag{4}$$

$$z_{ij} + (1 - z_{ik}) = 1 + w_{jk}, \; for \; i \in I, (j,k) \in V \tag{5}$$

$$(1 - z_{ij}) + z_{ik} = 1 + w_{jk}, \; for \; i \in I, (j,k) \in V \tag{6}$$

$$\sum_{i \in I} y_i \leq p \tag{7}$$

$$z_{ij} \in \{0,1\}, \; for \; i, \; j \in I \tag{8}$$

$$y_i \in \{0,1\}, \; for \; i \in I \tag{9}$$

$$w_{ij} \in \{0,1\}, \; for \; (i, \; j) \in V \tag{10}$$

The mathematical model for criterion average deviation can be written in the form:

$$Minimize \; dev_{avg} = \frac{\sum_{i \in I} \sum_{j \in J} \left| c_{ij} - n_{ij} \right| b_{ij}}{\sum_{i \in I} \sum_{j \in J} b_{ij}} \tag{11}$$

$$subject \; to \; (3) - (10)$$

Conditions (3) ensure that each station will be assigned exactly to only one zone. Conditions (4) ensure that the station *j* will be assigned only to the existing centre of the zone. Conditions (5) and (6) are coupling conditions between variables for allocation of the station to the zone and the variables for determining the zone border on the edge *(j,k).* Condition (7) ensures that we will create maximally *p* zones.

4 Solving technique

Both models will be solved using IP solver using the exact methods, so we will obtain exact solution of the problem. Because in model with criterion maximal deviation the objective function (2) is not a linear function, we need to modify this objective function to linear form. We introduce new variables u_{ij}, v_{ij} and x and reformulate mathematical model to the form:

$$Minimize \; dev_{\max} = x \tag{12}$$

$$subject \; to \; (3) - (10)$$

$$b_{ij} \left(c_{ij} - n_{ij} \right) = u_{ij} - v_{ij}, \; for \; i, \; j \in I \tag{13}$$

$$u_{ij} \leq x, \; for \; i, \; j \in I \tag{14}$$

$$v_{ij} \leq x, \; for \; i, \; j \in I \tag{15}$$

$$u_{ij} \geq 0, \; for \; i, \; j \in I \tag{16}$$

$$v_{ij} \geq 0, \; for \; i, \; j \in I \tag{17}$$

5 Numerical experiments

Verification of both models presented above was made on the data of the Zvolen County in Slovak republic. The stations in the network are represented by the 51 municipalities or parts of municipalities. In each node the main

station was selected to calculate the distance matrix and the OD matrix. Current prices are distance prices and were calculated according to real prices for travelling by regional buses in this area. The OD matrix was estimated using the quadratic gravity model [2], where the number of passengers between nodes i and j is calculated as follows:

$$\frac{b_i b_j}{d_{ij}}$$

Parameter b_i represents the number of inhabitants in the node i. Parameter p was set to 25.

To perform the computation we used the general optimization software tool XPRESS 7.1 [5]. This software system includes the branch-and-cut method and it also enables solving of large linear programming problems. 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 Table 2, Table 3 and Table 4 are results for both models used in this study. In the row "*f*" there are values of price for travelling one zone in the counting zones tariff system, in the row denoted as *Dev_max* are results for the model (12),(3)-(10),(13)-(17) and in the row denoted as *Dev_avg* are result for the model (11), (3)-(10).

In the Table 2 there are compared computational times of both models. In the Table 3 are compared the number of created zones and in the Table 4 there are calculated the total deviations between current prices and new prices for all passengers to compare the quality of solution.

f	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
Dev_max	0.484	0.484	0.593	0.671	0.765	0.827	2.418	1.996
Dev_avg	0.421	0.453	0.717	1.56	2.278	2.652	8.924	4.711
f	0.9	1	1.1	1.2	1.3	1.4	1.5	1.6
Dev_max	1.311	2.231	2.652	2.028	1.108	1.513	1.28	1.248
Dev_avg	5.398	6.069	9.017	2.465	2.948	2.402	1.482	1.482

Table 2 Computational times [in seconds]

f	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
Dev_max	25	25	25	25	25	25	25	25
Dev_avg	25	25	25	25	24	17	15	10
f	0.9	1	1.1	1.2	1.3	1.4	1.5	1.6
Dev_max	25	25	25	25	23	23	25	25
Dev_avg	11	8	6	5	5	3	4	3

Table 3 Number of created zones

f	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
Dev_max	27493.6	21023.8	15046.9	13450.3	13010.5	12649	15165.2	15403.7
Dev_avg	27493.6	20316.4	14184.4	12009	11386.5	9647.7	8582.1	8982.6
f	0.9	1	1.1	1.2	1.3	1.4	1.5	1.6
Dev_max	16249.6	18446.8	22036.7	25682.2	30029.6	34553.3	38781.4	43134.5
Dev_avg	9817.7	11010.6	12880.7	15086.5	17397.8	19653.9	21935.8	24522.6

Table 4 Total deviations between current prices and new prices for all passengers

6 Conclusion

The results of both models show that the computational time in both cases is acceptable, few seconds maximally. From the Table 2 we can see that in the average deviation model with increasing unit price f the number of created zones decrease. This is due to the different criterion in the objective function which takes into account all passengers in the system and not only the worst one, as it is in the case of maximum deviation model. From the

result in Table 3 we can see, that model with average deviation criterion gives better solution than model with maximal deviation criterion.

In the future we want to study also another different objective functions and compare this approach with other methods, for example heuristic methods mentioned in [1].

Acknowledgements

This work was supported by the research grants VEGA 1/0296/12 "Public Service Systems with Fair Access to Service", APVV-0760-11 "Designing of Fair Service Systems on Transportation Networks" and the institutional grant of Faculty of Management Science and Informatics (FRI ZU). We would like to thank to "Centre of excellence in computer sciences and knowledge systems (ITMS 26220120007) for built-up infrastructure, which was used.

References

[1] Babel. L., Kellerer, H.: Design of tariff zones in public transportation networks: theoretical results and heuristics, In: *Mathematical Methods of Operations Research* **58** (2003), Springer-Verlag, 359-374.
[2] Cenek, P., Janáček, J., Jánošíková, Ľ.: Location of transportation districts at modeling of tranportation processes in a region. In: *Communications - scientific letters of the University of Žilina* **4**, Issue 1-2 (2002), 5-9.
[3] Hajhalová, B.: Kedy sa stanú integrované dopravné systémy osobnej dopravy na Slovensku realitou? In: *Doprava a spoje : internetový časopis* **1** (2011), 29-36
[4] Schöbel, A.: *Optimization in Public Transportation: Stop Location, Delay Management and Tariff Zone Design in a Public Transportation Network*. Springer, 2006
[5] *XPRESS-Mosel User guide*. Dash Associates, Blisworth, 2005

Using technical analysis indicators in the terms of currency hedging

Roman Kolář[1].

Abstract Static hedging methods used just plain vanilla derivates and their combinations with no revision during hedged period. As hybrid hedging methods is considered using indicators of technical analysis for prediction of spot price in future and on the basis of this prediction it is decided between hedge company portfolio or stay unhedged.
For prediction are used these indicators: moving average, Relative Strength Index and Stochastic. Main goal of this work is to use technical analysis indicators for prediction currency pair value and on the basis of this prediction get positive part of risk. Effects of this strategy to cash flow of one hedged crown are tested on 11 years long time series of CZK/EUR currency pair. Starting days of historical testing are changed severa times due to providing better information. It helps us to find out hit ratio of this method. Further testing is advised.

JEL Classification: G11, C60
AMS Classification: 90C15

1 Introduction

Economics is a science where every rule every condition and every principle happens in the terms of probability. It means that we can not easily be sure about future. This elementary fact has to be realized by any decision making.

Events have their probability and are connected with risks. Risk is a deviation from expected situation which could be positive or negative. For example volatility of stock prices or commodity prices could have an effect (positive or negative) on cash flow of a company. It causes that medium and large companies have usually specialized department which is concentrated on risk management. Total risk can be divided between systematic and non-systematic risk. Non-systematic risk is connected with a specific asset and can be eliminated by diversification. Systematic risk is joined by whole market and only way how to lower it is a hedging. This text is focused on currency hedging using technical analysis indicators (e.g. Lo, Mamaysky, Wang [6]) as a tool for prediction which helps economics subjects to improve cash flows.

Using hedging eliminates negative and positive part of risk. Aim of this article is to quantify getting positive part of risk by using technical analysis indicators as moving average, Relative Strength Index and Stochastic. It will be tested on eleven years long time series of CZK/EUR currency pair.

Theoretical part is divided into two sections. Risks and basic classification is described in the first section. The rest of the first section is focused on hedging. Next section targets technical analysis. Described indicators are moving average. Relative Strength Index and Stochastic. Application of the technical analysis in currency hedging is evaluated in the last section. Achieved results are complied with a critic. Summary and implications are in conclusion.

2 Risks

Risk is a deviation of expected state which could be positive of negative. Corporations which make business in mining raw materials, manufacturing or providing services undergo risks joined by their entrepreneurial activity – e.g. Dubofsky and Miller [2].

Then there is a financial risk. This type of risk is connected with financial assets. Financial risks can have huge impact on company cash flows – e.g. Stulz [7]. The parting of the financial risks is in figure 1.

[1] VSB-Technical University of Ostrava, finance, Sokolská 33, Ostrava, R.Kolar@seznam.cz.

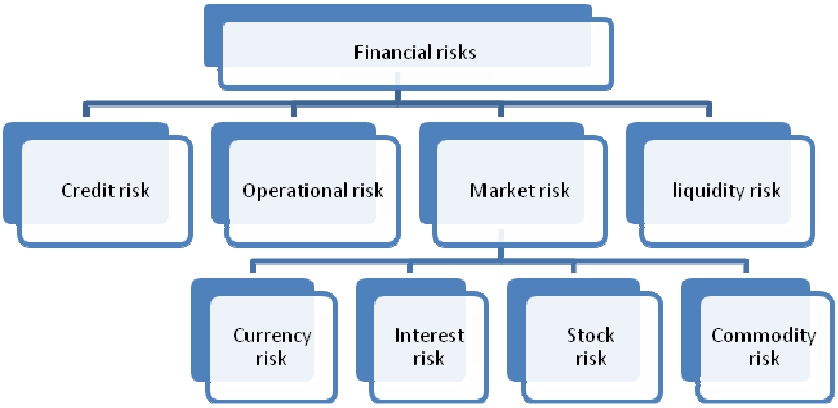


Figure 1 Classification of risks

Credit risk is caused by default of other side of agreement or contract. Operational risk is joined with failure of system or human factor. Market risk is connected with specific type of asset as it is seen in the Figure 1. At the end there is liquidity risk which is caused by not having enough current assets to pay debts or other liabilities. By the way of elimination there are systematic and non-systematic risks which are described in introduction.

There are many types of risks as it is seen in Figure 1. In practical part of this text will be described method using chosen technical analysis indicators for getting positive part of chosen risk in the terms of hedging. It is necessary to describe hedging and its basic classification.

3 Hedging

Key point about hedging is creation of a portfolio which consists of risk asset and one or more other assets with inverse payoff function. Hedgers try to eliminate influence of volatility of the risk asset to value of portfolio – e. g. Stulz [7]. Value of the hedge portfolio can be evaluated by followed equation:

$$\Pi_t = Q \cdot S_t - h \cdot N \cdot f_{t,T} \tag{1}$$

Π_t is value of the portfolio. S_t s price of risk asset at time "t" and $f_{t,T}$ is price of financial derivate used for hedging. Q represents quantity of risk asset and N represents quantity of underlying asset per unit of financial derivate. Last symbol is h which represents hedge ratio.

There are many ways how to class hedging strategies. Classification can be following:

According to hedging level:

- perfect hedging,
- super hedging,
- partial hedging,
- unhedged portfolio.

According to frequencies of revisions:

- continues,
- discrete.

According to type of eliminating risk:

- total risk,
- systematic risk,
- market risk,
- credit risk,
- operational risk.

According to number of revision during time:

- static hedging,
- dynamic hedging.

Corporations can use internal methods of hedging or external methods of hedging. AS the internal hedging methods can be considered netting, matching, lagging, leading, monetary diversification and set up of agree-

ments. External methods of hedging use financial derivates to close up risky position. Financial derivates are wide group of assets which can be specified as agreements between two parties with different number of conditions vide Hull [4].

There are many ways how companies can hedge against different type of risk. Technical analysis indicators give corporations information about future spot price o asset with some probability. Used indicators are described in next section.

4 Technical analysis

Using technical analysis for predicting development of assets price is known for a long time. There are a lot of books and articles about using or testing it – e. g. Elder [3]; Blume, Lawrence, David Easley and Maureen O'Hara [1]; or Lo, Mamaysky, Wang [6] etc. .

In practical part of the text are described methods using indicators of technical analysis for a prediction. On the basis of the prediction company decides between staying unhedged or be hedged.

There are two quite huge approaches how to analyze charts by technical analysis. The first one are technical formations in charts as double bottom, support/resistance, trend lines etc. Conception of technical formations is too subjective. On the other hand there are indicators which are calculated by more or less difficult algorithms vide Elder [3]. There are used these indicators: moving average. Relative Strength Index and Stochastic.

Moving averages (further also MA) are used by statistic science for a long time. This indicator displays directly into price chart. Formula of simple moving average is following:

$$SMA(N) = \frac{p_t + p_{t-1} + \ldots + p_{t-N}}{N} \tag{2}$$

where N is period of moving average and P_t is symbol for price of assets.

Relative Strength Index (further RSI) is the indicator which develops into their oven secondary chart. Curve of RSI oscillates between 0 and 100. The first step for calculation is to count upward change and downward change. There is a formula for upward change:

$$U = P_t - P_{t-1}; D = 0 \tag{3}$$

Where U is absolute upward change and D is absolute downward change which is 0. In the same analogy is counted downward change which formula is followed:

$$D = P_{t-1} - P_t; U = 0 \tag{4}$$

Symbols are the same. Next step is to count Relative Strength by this formula:

$$RS = \frac{MA(U,N)}{MA(D,N)} \tag{5}$$

Where $MA(U,N)$ is moving average upward changes during the period and $MA(D,N)$ is moving average of downward change during the period. RSI is counted by this formula:

$$RSI = 100 - \frac{100}{1 - RS} \tag{6}$$

Stochastic also oscillates in secondary chart between 0 and 100 but way how to calculate them is different. This indicator uses two curves (%K and %D). %K curve also known as pure Stochastic can be calculated by this formula:

$$\%K = \frac{C_t - P_{low,n}}{P_{high,n} - P_{low,n}} \tag{7}$$

where C_t is today close price, $P_{low,n}$ is the lowest price during the period and $P_{high,n}$ is the highest price during the period. %D Stochastic is calculated as an exponential mowing average of %K Stochastic with period 3 vides Elder [3].

5 Using of indicators of technical analysis in the terms of hedging

Hedging is about effort to minimize risks but indicators of technical analysis can be used for prediction of future price of any assets. On the basis of the prediction company can decide about be hedged or staying unhedged (trying to get positive part of risk). Let´s calculate how useful the described indicators can be.

We have to mention some general information about hedged position at the beginning of this section. Total amount of an asset was 60 300 000 EUR. For short hedge were used these financial derivates: short forward, put option. There were also used these option spreads: long strangle, long straddle, strip, strap. Using the short forward for hedging contributed to stabilized cash flows. On the other hand using put options and options spreads leads to higher volatility of cash flow with some probability to reach gain. All further information is in Kolar [5]. Results of using technical analysis indicators will be calculated in Crowns per one hedged Euro. Whole procedure of testing is described in following paragraphs.

There are many technical analysis indicators and described ones before are just small sample. Each of these indicators instead of moving average oscillates between 2 extreme values and it causes that there are one exact middle value (between extremes) which is used as a border. If the value of indicator is below this border, it will be considered as a signal for decline of CZK / EUR and vice versa. Moving average does not oscillate in its own chart and this is why the prediction is a little bit different. If the moving average is below current parity of CZK / EUR, it will be considered as a signal for a rise and vice versa. RSI and both Stochastics are computed in their origin settings. The best Settings of moving average have to be found out.

For testing of indicators' hit rate is used historical time series CZK / EUR from 1 January 2000 to 31 December 2011 (daily data). Indicators' values will be calculated for each day of 11 years long time series. In order to above mentioned scheme they will be used for prediction. We suppose that value of indicator today predicts value of parity in 21 working days. For example suppose that starting day is on 31 March 2000 and indicators' values are used for prediction of the parity CZK / EUR on 2 May 2000 (period lasts 21 working days). Starting day will be several times changed which gives us better information about this strategy.

Now we know how to use indicator for prediction. But there is no specification of company. In next section is supposed Czech company which has receivable in Euros. If the indicator predicts rise in parity CZK / EUR, company stay unhedged to proof cash flows and vice versa. There is no need for quantification of monetary exposition (results will be calculated in CZK per on hedged Euro). Used hedging method in case of predicting decrease of the parity is also not important because this article is focused on using technical analysis indicators and their contribution to company's cash flow.

According to above mentioned assumption were calculated cash flows of hedging strategies using indicators of technical analysis. Starting day is 31 March 2000. Results are in Table 1 (results are in crowns for one hedged euro). During eleven years long time series there are 149 observations. For better information about using indicators we can change starting day a find out how results will change.

Year	Non-hedged	MA	RSI	%K	%D
2000	-0.9450	-1.0850	-0.2550	-0.1000	-0.1700
2001	-1.5400	-1.2700	-1.1050	0.0000	0.6150
2002	-1.9100	0.1800	-1.0900	1.5500	1.4800
2003	1.3300	1.1250	1.0000	0.7400	1.3250
2004	-1.7750	-0.3150	-0.0550	0.6300	0.3600
2005	-1.8300	-1.1050	-0.1650	-0.2600	0.2200
2006	-0.9850	-0.7250	0.1350	-0.4450	0.1700
2007	-1.9300	-0.2450	-0.2450	0.7750	0.7750
2008	0.2750	1.0200	1.1800	0.2250	0.2500
2009	-0.5650	0.2150	0.2100	-1.0850	0.5400
2010	-0.4950	-1.0150	0.2600	1.0150	1.1200
2011	0.5450	0.2450	-0.2450	0.9450	0.7050
Sum	**-8.8800**	**-1.8900**	**-0.1200**	**4.0900**	**7.5600**

Table 1 Yearly comparison of indicators

We can see that only both Stochastic indicators could make profit in total sum. In comparison with non-hedged approach were results of all indicators much better. On the basis of data from Table 1 is the best %D Stochastic which increases cash flows in almost each year.

For better information about using indicators we can change starting day and find out how results will change. Total sum of each indicator is calculated in Table 2 for different starting days.

Starting day	Non-hedged	MA	RSI	%K	%D
31. 3. 2000	-8.8800	-1.8900	-0.1200	4.0900	7.5600
4. 4. 2000	-8.8950	-1.2200	-2.5900	2.9200	3.0450
6. 4. 2000	-9.1000	-1.0500	-3.8000	3.4700	2.6150
10. 4. 2000	-9.2900	-3.5200	-3.0060	5.0350	6.5810
12. 4. 2000	-9.3600	-2.9000	-0.0300	8.1350	8.9500
14. 4. 2000	-9.3800	-1.3350	0.9600	6.8050	6.7300

Table 2 Influence of different starting day to total sum of earnings

Starting days were changed by 2 working days as you can see in the Table 2. The best results still keeps stochastic indicator but we can see high volatility of results. High volatility of total sums improves that using indicators of the technical analysis can raise cash flows at the cost of higher risks. If we have a look at the best tested indicator which is %D Stochastic we can figure out mean about 5.9135 Crowns per one hedged Euro. Standard deviation is 2.3149 Crowns per one hedged Euro. All indicators reached better results than non-hedged approach but just Stochastic indicator could reach gain which is very variable as we can see at Table 2.

6 Conclusion

In the first section is described classification of risks. The second section is focused on hedging and its classification. The end of theoretical part of text is the third section which is concentrated on technical analysis. Practical part of this text is represented by the fourth section.

The target of this work is to quantify of using technical analysis indicators by currency hedging. Used indicators are moving average, Relative Strength Index, %K Stochastic and %D Stochastic. Results of the first observation with starting day in 30 March 2000 were quite good. The best results reached Stochastic with total gain 7.56 crowns per one hedged euro. Testing on historical time series gives us 149 possibilities for hedging. On the basis of Stochastic signals there were 47 unhedged periods. 14 signals were false and caused loss. Changing of starting day causes high volatility of results. Both Stochastic indicators reached gain in total sum. %K Stochastic reached average 5.0758 crowns per 1 hedged euro with standard deviation 1.8506 crowns per 1 hedged euro. %D stochastic reached average total sum and standard deviation 5.9135 and 2.3149 crowns per 1 hedged euro. Changing starting day just about several days can cause high volatility of outcomes and that is why using of this approach for hedging is controversial.

Companies with risk-averse management have higher total utility due to elimination of total risk but there are also risk seeking subjects which can try to undergo higher risk with a possibility of higher profit and for these companies can be using indicators of technical analysis more interesting. It was shown that there is higher profit potential but the price for it is higher risk. Each company using hedging can use this approach because of no transaction costs and high variability of risky assets which price can be predicted by technical analysis indicators. All that matters is relation to risk of the subject.

Acknowledgements

The support provided by VSB-TU Ostrava under the SGS project SP2012/2 is kindly announced.

References

[1] Blume, L., Easley, D., and O´Hara, M.: *Market statistics and technical analysis: The role of volume. Journal of Finance* 1994. ISSN 153–181.

[2] Dubofsky, D. A., and Miller, T. W.; *Derivates, Valuation and Risk Management*. 1st ed. New York: Oxford University Press. 2003. 646 p. ISBN 0-19-511470-1.

[3] Elder, A.:*Trading for living*. 1st ed. Wiley. 1993. 289 p. ISBN 978-0471592242.

[4] Hull, J. C.; *Option. Futures and Other Derivatives*. 6th ed. New Jersey: Prentice Hall. 2005. 744 p. ISBN 0-13-149908-4.

[5] Kolar, R.: *Zajištění měnového a komoditního rizika vybraného podniku.* VŠB-TU Ostrava. 2012.

[6] Lo., Mamaysky., and Wang.: *Foundation of Technical Analysis: Computational Algorithms. Statistical Inference and Empirical Implementation. Journal of Finance*. 2000. ISSN 42-111.

[7] Stulz, R. M. *Risk Management & Derivatives*, 1st ed. Mason: Thomson. 2003. 676 p. ISBN 0-538-86101-0.

[8] Tichý, T.: *Posouzení metody částečného hedgingu na případu řízení měnového rizika nefinanční instituce. Ekonomická revue – Central European Review of Economic Issues*. 2009. Vol. 12. pp. 69-82. ISSN 1212-3951.

[9] Zmeškal, Z.: *Finanční modely*. Ekopress. 2004. 236p. ISBN 80-86119-87-4.

Implementation of blending problem algorithm into information system

Pavel Kolman[1]

Abstract. Optimal feed mix is very important part of cattle nutrition problem. For filling in of missing ingredients are used food addictions. They are produced according to needs of concrete breeder. All the ingredients in animal feed addictions must occur in exactly given amount. This requirement is determinative for breeder, for feed addictions producer are most important minimal costs. To find optimal ratio of nutrients to get all ingredients in required amount (or in given bound), it is used blending problem. It belongs to linear programming and is easily solvable with simplex method.

This paper describes problematic of blending problem implementation, when finding optimal nutrition ingredients ratio in producing company. Requirement of company is to have blending problem (include simplex method algorithm, solving this problem) implemented in their information system, so that they were able to optimize nutrient ratio of each food addiction they produce.

Keywords: Nutrition problem, Blending problem, Simplex method, cycling in the Simplex method, Delphi, Programming.

JEL Classification: C44, C61
AMS Classification: 46N10, 90C05, 90C08

1 Introduction

Costs minimization is a very important factor for company which is blending feed addictions. That is the main reason, why there is used optimization in a decision process, which nutrients and in what amount will be used for food addictions mixing process. Character of solved problem allows creating mathematical model solvable with linear programming methods. For optimization it is used the simplex method. In this paper there will be described problematic of simplex method algorithm implementation into information system. This information system is being developed for producing company. The concrete name of developing company, nor producing company will not be mentioned, from competitive reasons.

2 Material and methods

When finding optimal composition of food addictions, in the first step it will be necessary to make a mathematical model of described problem. The variables in mathematical model will be searched amounts of nutrients in given units (they will differ according to used nutrient). Constraints in mathematical model will describe amounts of ingredients (minerals, proteins etc.) in minimal, maximal amount or given bound. With regard to mathematical model character of blending problem there will be always used simplex method.

The simplex method will be applied according to generally known rules, with exception to pivot selection. Because the simplex algorithm is very well known, it will be described very briefly. In the first step, mathematical model will be converted into standard form, i.e. all the inequalities will be changed to equations by adding additional variables (slack or surplus). Then follows making of canonical form of LP problem, where there will added artificial variables so that there was identity submatrix. This canonical form is the initial basic solution. The next steps consist in individual iterations, considering from optimality test, pivot selection and basis change. While for optimality test and basis change will be used well known rule as described in literature, pivot selection rule will differ.

Reason for different pivot change rule is elimination of potential cycling in solved mathematical model. It can occur only in degenerated problems, but this is a quite often phenomenon in blending problem models. As a cycling we understand situation, when after several iterations in simplex algorithm we get to the same basis, we have already been before. As a necessary condition for cycling occurrence is zero objective function change, because its value within individual iterations cannot worse. If there would be used generally known pivot selection rule according to Jablonský [4] or Stevenson [5], some of problems (e.g. Todd [2]) would get into a cycle. For cycle elimination there will be used Blend's pivot selection rule.

Blend's pivot selection rule chooses as a variable entering the basis that with its minimal index. The algorithm describes e.g. Blend [1].

[1] Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operation Analysis, Zemědělská 1, 60300 Brno, xkolman@mendelu.cz.

When the application has been developed, for pivot selection there was used the combination of both approaches. Because cycling can occur only in situation, when objective function change is 0, the newly defined pivot selection algorithm was following:

- Pivot was selected with according to the original simplex algorithm. When its selection has improved objective function, was used for basis change.
- When the objective function improvement was zero, pivot was chosen according to Blend's algorithm.

It is obvious, that in this modification cannot occur cycling, because in all potential cycle occurrences it is used Blend's anti-cycling rule which excludes this eventuality. Textbook example of LP problem where can occur cycling is visible on Figure 2**,** describing input data format. It was gained from Todd's [2] demonstration of cycling simplex method. For terminating of algorithm calculation is sufficient optimality criterion fulfillment. If obtained solution is feasible or nor will be decided considering to blending problem specifics directly in information system regarding to structural variables values.

3 Results

Before description of proper program and its functions, it should be mentioned reason why the program was developed in Delphi as a console application. The information system itself is being developed in Microsoft Dynamics NAV, also known as Navision. Navision is not suitable for large problems calculations, but this is typical property of blending problem models. From this reason it was agreed, that communication interface will be solved in Navision and own calculations will be realized by console application. The application will be available to information system in exactly specified directory and information system will be allowed to run it always when necessary. Advantage of this approach is a fact that whole console application takes after compilation only 100 Kbytes of memory! The reason of such small size is mainly caused by absence of graphical interface in application.

3.1 Program structure

The own program, that will be executed from information system, will be console application programmed in Delphi. For own calculations will be run with 2 parameters. The parameters itself will be input file with mathematical model in required format (this model will be information system output and console application input) and output file will be optimal solution of LP problem (this file will be the other way around console application output and information system input). Because the program is made as a console application, it can be run parametrically from command line, or directly from some programs. This is very important requirement, mainly from reason, that information system will run the application automatically, without user notification. The execution of program with two parameters will be following:

```
simplexka.exe parameter1 parameter2,
```

where `simplexka.exe` is program name, parameter1 contains name of file with mathematical model (alternatively with path) and parameter2 is name of file, where the optimal solution will be saved. After execution with parameters the mathematical model is loaded, solved and optimal solution saved to file from parameter2. Here should be mentioned 2 facts. First of all, if file from parameter2 exists, is automatically (without warning) overwritten by newly created file. Second of all, if program has no access rights to directory with optimal solution file, the file with optimal solution will not be saved. For verification, if solved problem has feasible solution or no, is intended program exitcode. In case that optimal solution is found, the exitcode value is set as 0, otherwise 1. Verification, which requirements has not been fulfilled and if those breakings are important or no (e.g. if non fulfillment of constraint is in thousandths of percent, constraint can be although considered as fulfilled) is depending on information system developers.

The main application source code is displayed in Figure 1 and is relatively short. The main reason is that all functions needed for successful problem solving are included in `simplexka.pas` unit. This unit (library) is a necessary part of program and was programmed to be usable in other applications where simplex method is needed. Unit simplexka.pas detailed description transcends range of this paper, therefore will not be furthermore described. So what happens, when the program is executed? First of all, the parameters are load. The input file is saved to input variable, output file to output variable. Then the mathematical model saved in input file is load to LP variable. It is structured data type containing whole linear programming problem. Afterwards, the

`vyresit_ulohu_LP` function is run. The function has 2 parameters: the first parameter is mathematical model contained in LP variable, the second is LP problem solution vector. It is a dynamic data field of real48 data type. After its termination the function returns Boolean value true in case, that optimal solution of mathematical model in LP variable was found. If obtained solution is not feasible, return value of function is false. According to function return value, application exitcode is assigned. The last function of main program is `zapis_reseni`

function with 2 parameters. The first parameter is dynamic data field containing optimal values of structural variables the second one is file, where this optimal solution will be saved. The main application source code is visible on Figure 1.

```
program krmiva;
{$APPTYPE CONSOLE}
uses
  SysUtils,
  simplexka in 'simplexka.pas';
var
  input, output: string; { input and output file }
  lp: TLPproblem; { LP model }
  reseni: Treseni; { LP model solution }
begin
  try
    { load input and output file from parameters }
    input := ParamStr(1);
    output := ParamStr(2);
    { loads model to LP variable }
    lp := nacti_ulohu(input);
    { solves LP problem and results if solution is feasible or no }
    if vyresit_ulohu_LP(lp, reseni) then
      exitcode := 0
    else
      exitcode := 1;
    { writes solution to file }
    zapis_reseni(output, reseni);
  except
    on E: Exception do
      Writeln(E.ClassName, ': ', E.Message);
  end;
end.
```

Figure 1 The main application source code.

3.2 Input data

Because the information system is being developed in Navision and application solving the blending problem is executable *.exe file, it was necessary to agree on communication between information system and program. With information system developers it was agreed, that mathematical model will be put together in information system in a specific format, and saved to exactly specified place. Therefrom it will be load from application, solved and saved to previously specified place. At the same time, the problem solution will be load from information system and processed. Mathematical model construction will be implemented directly in information system. Its correct construction will handle person responsible for food addictions blending. The main application task is correct solving of LP problem and return of result in specified format to information system. Because the program works in Czech computer environment, the decimal separator is comma. Example of mathematical model in specified format (in detail will be described below) is on Figure 2.

```
[Objective function]¶
Z MAX¶
10 → -57 → -9 → -24¶
[Binding constraints]¶
0,5 → -5,5 → -2,5 → 9 → <= → 0¶
0,5 → -1,5 → -0,5 → 1 → <= → 0¶
1 → → → → <= → 1¶
```

Figure 2 Example of cycling LP problem in input file format.

The first row contains comment, that objective function follows and has only informative meaning for user. It can contain any text and during import is only loaded, but not processed. From second row, the LP type is loaded. Although it is assumed that there will be solved only minimization problems (resulting from blending problem properties), regarding to program universality there remained this option. When the solved problem is maximization type, text string contains "Z_MAX" value, for minimization problems "Z_MIN". On the third row there

are objective function coefficients separated by tab sign (#9 in ASCII code) and ended by end of line sign (CR-LF in Windows operation system, in ASCII code represented by #13#10). On the fourth row, there begins loading of constraints. This row has only informative meaning, the same as first row. Since fifth row, the constraints are founded, each constraint on one row. The structural coefficients separator is in the same way as in objective function the tab sign. The constraint is in form structural coefficients, relation and right-hand side, all separated by tabulators. Example of mathematical model in described format is visible on Figure 1. The main advantage of previously described format is its universality. It is very easy to take any mathematical model and in any text editor or table processor, prepare it to given format.

In this situation here should be mentioned case, where a structural or objective function coefficient is zero valued. Regarding easy transport from or to MS excel (mainly through clipboard), it is possible instead of zero values write empty strings (separated with tabulators). Then, zero valued coefficients are recognized according to two consecutive tabulators, e.g. see last row of Figure 2 example, where there are 3 zero valued structural coefficients.

The question of correct mathematical model loading is very important for practical use of program. It is not acceptable to load wrong any constraint or, not to load it. Therefore there should be mentioned, how does the application identifies real number of variables and constraints in a model. When the mathematical model is being loaded, the first row is ignored. Number of structural variables is specified from 4^{th} non-empty row. It is the row, where first constraint occurs. Number of structural variables is specified as number of tab (#9) signs before first occurrence of "<" or "=" or ">" sign. Number of constraints corresponds to number of non-empty rows, decreased by four. From these four rows, first three are related to objective function and fourth row informs user that from following row there are individual constraints. Although this format is not absolutely resistant to wrong model loading, when there is used not correct data format, after consultations with information system developers was considered as sufficient. Because the mathematical model construction will be in information system realized automatically, the probability of making such model is extremely low. This problem can occur (with relatively high probability), when the mathematical model is constructed manually by user.

3.3 Output data format

Regarding to a fact, that program output (i.e. optimal solution) will be processed in information system, it was necessary to agree with developers on output format. It is very simple: optimal solution values of all structural variables (include non-basic variables, i.e. zero valued variables) are written to output file, each variable on one row with precision of 8 decimal positions. Mathematical data format is not allowed. This format guarantees easy and correct loading of optimal solution into information system.

4 Discussion

Development of the application is up to date finished. With cooperation with information system developers, the small error eliminations and improvements are under way. During the developing process, the great accent was focused out on absolute possible results precision and elimination of cycling problem.

The maximal precision of results was solved by using of real48 data type. According to Delphi-basics [3], it is the floating point type with the highest capacity and precision.

The more difficult problem to solve was elimination of potential cycling. In a practical use, this could cause serious problems. This problem was solved by alternative way of pivot selection. It was used combination of original and Blend's rule. When there was used standard pivot selection and the solved problem was degenerate, the algorithm cycled relatively often. By small change (when objective function change is 0, we choose pivot column breaking optimality criterion with lowest index) was occurrence of cycling eliminated. The program itself was tested on examples, where there is a high probability of cycling: assignment problems solved as LP problems by simplex method. Those problems are strongly degenerate and therefore here very often occur cycling. During testing the program was solving of different assignment problems up to 30x30 dimensions (reason, why there were not tested larger problems consists from fact, that solving time grew exponentially with growing dimensions of matrix). While in original way of pivot selection the cycles occurred often in 7x7 matrix and program had to be terminated manually, improved version never cycled.

The last serious problem, that has not been solved yet, but will have to be solved together with information system developers and blending company employees, is infeasibility tolerance. In blending problem regularly occurs situation, that optimal solution of given blending problem does not exist. But when one or more of the constraints is softened, problem has feasible solution. The plan for future is to let program automatically decide, when small breach of any constraint is yet "all right" and when no. Now, the decision process fully depends on program user.

5 Conclusion

Although the program development has not been finished, the project state goes to finish and is program is prepared to be used in information system. It is possible that in practical use there occur problems with program they will have to be solved. But, those problems are not known yet. During testing the program works stable and program results correspond with results obtained from commercial optimization software. But, for real practical experience it will be necessary to wait.

References

[1] Blend, R. J.: *New finite pivoting rules for the Simplex method*. Mathematics in operations research, Binghamton, 1977. Available on http://sci.sut.ac.ir/People%5CCourses%5C42%5Cbland-rule.pdf.

[2] Todd M. J.: *Cycling in the Simplex method*. Available on http://people.orie.cornell.edu/~miketodd/or630/SimplexCyclingExample.pdf

[3] *Delphi basics*. Delphi manual, available on http://www.delphibasics.co.uk/RTL.asp?Name=Real48

[4] Jablonský, J. *Operační výzkum : kvantitativní modely pro ekonomické rozhodování.* 3. vyd. Praha: Professional Publishing, 2007. 323 s. ISBN 978-80-86946-44-3.

[5] Stevenson, W. J. -- Ozgur, C. Introduction to management science with spreadsheets. Boston: McGraw-Hill/Irwin, 2007. 812 s. ISBN 978-0-07-325290-2.

Robustness and bootstrap approaches to SSD portfolio efficiency testing

Miloš Kopa[1]

Abstract. This paper deals with portfolio efficiency testing with respect to second-order stochastic dominance (SSD) criteria. Unlike the pair-wise tests the portfolio efficiency tests allow for full diversification across the assets. As usual in SSD testing, the returns of assets are modeled by scenarios. We apply a computationally attractive method to test whether a US market portfolio, proxied by CRSP all share index, is SSD efficient with respect to 48 US industry representative portfolios. Moreover, we present a robust version of SSD portfolio efficiency test that allows for small errors in data and we analyze their impact on the market portfolio SSD efficiency. We enrich the results by the estimation of p-value of market portfolio SSD efficiency when the bootstrap technique to the data is applied.

Keywords: Portfolio efficiency, second-order stochastic dominance, robustness, bootstrapping

JEL classification: D81, G11
AMS classification: 91B16, 91B30

1 Introduction

Portfolio efficiency tests with respect to second-order stochastic dominance (SSD) present a powerful non-parametric tool of decision-making theory for verifying the admissibility of a given portfolio for risk averse investors. The basics of decision-making theory were presented in the seminal work of Harry Markowitz [12]. He identified two main components of portfolio performance, mean reward and risk represented by variance. Applying a simple parametric optimization model he found the optimal trade-off between these two components. In this case, the portfolio is classified as efficient if there is no better portfolio, i.e., a portfolio with a higher mean and smaller variance. In the last 60 years, the theory of mean-risk models has been enriched by other risk measures, for example, semivariance, see [13], Value at Risk (VaR) or Conditional Value at Risk (CVaR), see [15], [17], [18].

More advanced application of risk measures in portfolio efficiency was introduced in Data Envelopment Analysis (DEA) models. Recently, Branda and Kopa [1] formulated DEA-risk models with risk measures as inputs and mean gross return as the output. These DEA-risk models can be seen as a generalization of mean-risk models, because they allow for multiple risk measure application. Moreover, if only one input is considered, then DEA-risk efficiency implies mean-risk efficiency with respect to the same risk measure. Branda and Kopa [1] compare the results of DEA-risk models with those of SSD portfolio efficiency.

Alternatively, one can adopt utility functions [14] for modelling investor's risk attitude, especially in the maximising expected utility approach. If the utility function is perfectly known, one can find the optimal decision. If that is not the case, one can at least identify the set of efficient portfolios with respect to a chosen class of utility functions. Considering all utility functions, that is, assuming only non-satiation for the investor's preferences, leads to the first-order stochastic dominance (FSD) relation (see [11] and references therein). Usually, we assume that the decision maker is risk averse, what reduces the considered class of utility functions. The admissible utility functions for risk averse investors are the concave ones. Therefore, adding the risk aversion assumption leads to the second-order stochastic dominance rules.

Using pairwise comparisons (e.g. [6]), an alternative (asset) is classified as SSD efficient if there is

[1]Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Praha 8, Czech Republic, kopa@karlin.mff.cuni.cz

no other alternative that dominates the former alternative with respect to SSD. It means that a given portfolio is SSD pairwise inefficient if there exists some asset that is preferred by all decision makers. This is often very strong condition and even if it is violated still the alternative may not be the optimal choice for any investor. Therefore, Fishburn [5] defined the convex stochastic dominance efficiency as follows: an alternative is convex SSD inefficient (dominated by other alternatives) if every investor prefers at least one other alternative. Despite that, it does not cover the full diversification case if investors may combine assets in portfolios. Tests for SSD portfolio efficiency allowing full diversification across the assets were developed in [16], [10], and [8]. These tests classify a given portfolio as SSD portfolio efficient if there is no portfolio created from the assets that SSD dominates the portfolio. Alternatively, one can also apply FSD efficiency tests, see [10] and [9].

In all SSD efficiency tests a scenario approach to asset's returns is considered. Unfortunately, the results of these tests are very sensitive to changes in scenarios. Even a small perturbation of data matrix can completely change the SSD classification of a given portfolio, that is, a portfolio that was originally classified as SSD efficient turns to be SSD inefficient for perturbed data. Therefore, Kopa [7] suggests robust versions of SSD efficiency test based on δ-SSD portfolio efficiency. In these tests, small perturbations of the original scenarios are allowed and moreover, maximal possible changes in values of scenarios that do not change SSD classification are identified. Alternatively, one can adopt robustness approach of [4] that uses contamination techniques (recently applied also in [2]). Given the alternative distribution (for example stress scenario) on can explore the stability of SSD classification with respect to contaminated distributions. Dupačová and Kopa [4] introduced directional SSD portfolio efficiency approach that classifies portfolio as directionally SSD efficient (inefficient) if it is SSD efficient (inefficient) when using the original data as well as when the data are contaminated by an alternative distribution. They also derived conditions that are necessary or sufficient for directional SSD (in)efficiency.

Another way of dealing with high scenario sensitivity of SSD efficiency tests was proposed in [16] and [9]. They applied bootstrap techniques and evaluated the bootstrap p-value of market portfolio efficiency. They generated 10 000 pseudo-samples. In each pseudo-sample they tested portfolio efficiency of US market portfolio with respect to second-order [16] and first-order [9] stochastic dominance criteria. Their results reject FSD (SSD) portfolio efficiency with very high reliability.

In this paper we employ both robustness and bootstrap techniques and we compare the information obtained from results of both approaches. Contrary to [7], we introduce robust versions of the Post test instead of the modified Kuosmanen test. These new tests are more computationally attractive and allow for changes only in returns of market portfolio. Since returns of US market portfolio are proxied by (CRSP) all share index, we identify the maximal possible error in return scenarios of the market portfolio to preserve its SSD (in)efficiency. In bootstrap application, we follow [16] and [9] but we use new data set including during crises scenarios and we consider different representative (industry) portfolios as the basic assets.

The remainder of this paper is structured as follows. Section 2 presents notation, basic definitions and recalls SSD portfolio efficiency tests. It is followed by introduction of new tests, that can be seen as robust versions of the Post test. Section 4 shows the basic data description and presents the results of both tests as well as of the bootstrap approach. The paper is summarized and concluded in Section 5.

2 SSD portfolio efficiency tests

We consider a random vector $\mathbf{r} = (r_1, r_2, \ldots, r_N)$ of returns of N assets with a discrete probability distribution 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 $\mathbf{x}^t = (x_1^t, x_2^t, \ldots, x_N^t)$ is the t-th row of matrix X representing the assets returns along t-th scenario. We assume that the decision maker may also combine the alternatives into a portfolio. We will use $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_N)^{\mathrm{T}}$ for a vector of portfolio weights and $X\boldsymbol{\lambda}$ represents returns of portfolio $\boldsymbol{\lambda}$. The

portfolio possibilities are given by a simplex

$$\Lambda = \{\boldsymbol{\lambda} \in R^N | \mathbf{1}'\boldsymbol{\lambda} = 1, \ \lambda_j \geq 0, \ j = 1, 2, \ldots, N\},$$

which arises as the relevant case if we exclude short sales and impose a budget restriction. Moreover, the random return of the tested asset $\boldsymbol{\tau}$ is denoted by r_{N+1}. It takes values y^t, $t = 1, ..., T$ with equal probabilities. Without loss of generality, we assume that scenarios $\mathbf{x}^t$ are ascendantly ordered according to returns of the tested asset, that is, $y^1 \leq y^2 \leq ... \leq y^T$.

Following [11] and references therein, portfolio $\boldsymbol{\lambda}$ *dominates* asset $\boldsymbol{\tau}$ *with respect to second-order stochastic dominance* ($\boldsymbol{\lambda} \succ_{SSD} \boldsymbol{\tau}$) if $Eu(\mathbf{r}\boldsymbol{\lambda}) \geq Eu(r_{N+1})$ for all non-decreasing and concave utility functions with strict inequality for at least one such utility function. Alternatively, one can consider as a definition of this relation some of its necessary and sufficient conditions summarized in, for example, [7]. In any case, if portfolio $\boldsymbol{\lambda}$ dominates asset $\boldsymbol{\tau}$ with respect to second-order stochastic dominance then every risk averse decision maker prefers $\boldsymbol{\lambda}$ to $\boldsymbol{\tau}$ or is indifferent between them.

Following [16], we define the efficiency of a given portfolio with respect to second-order stochastic dominance relative to all portfolios that can be created from a considered set of assets.

Definition 1. A given asset $\boldsymbol{\tau}$ is *SSD efficient* if there exists a concave utility function u such that:

$$Eu(r_{N+1}) > Eu(\mathbf{r}\boldsymbol{\lambda}) \ \ \forall \boldsymbol{\lambda} \in \Lambda. \tag{1}$$

Otherwise, asset $\boldsymbol{\tau}$ is *SSD inefficient*.

Since Post [16] proved that the representative set of utility functions for SSD efficiency consists only of piece-wise linear concave functions, one can easily search for the admissible function (that is, function satisfying (1)) using linear programming techniques. Summarizing, if there is no tie in scenarios of tested asset, Post [16] derived the following SSD efficiency test.

Theorem 1. *Let*

$$\theta^* = \min_{\theta, \beta_t} \theta \tag{2}$$

$$\begin{aligned} s.t. \qquad \sum_{t=1}^{T} \beta_t (y^t - x_n^t) + T\theta &\geq 0 \quad n = 1, 2, ..., N \\ \beta_t - \beta_{t+1} &\geq 0 \quad t = 1, 2, ..., T-1 \\ \beta_t &\geq 0 \quad t = 1, 2, ..., T-1 \\ \beta_T &= 1. \end{aligned}$$

A given asset $\boldsymbol{\tau}$ is SSD efficient if and only if $\theta^* \leq 0$.

If the tested asset is SSD efficient then an admissible utility function can be constructed from marginal utility levels β_t^* identified by (2) as the optimal solutions. On the other hand, the optimal coefficients β_t^* have no economic meaning for SSD inefficient portfolio.

Beside the Post test, one can alternatively use the Kuosmanen test [10] or the Kopa and Chovanec test [8] for SSD efficiency of a given asset or portfolio. Both these tests are more computationally demanding than the Post test, however they provide a useful information about a SSD dominating portfolio if the tested one is SSD inefficient.

The optimal value of the Post test (as well as of the Kuosmanen test and the Kopa and Chovanec test) can be seen as a degree of inefficiency. However, if the tested portfolio is SSD efficient these tests give no additional information. Therefore, Kopa [7] introduced a measure of SSD efficiency that is applicable for SSD efficient portfolios.

3 Robust versions of the Post SSD efficiency test

Since SSD efficiency of a given asset is very sensitive to any changes in scenario vector $\mathbf{y}$, we present robust versions of the Post test, that classify a given asset as SSD (in)efficient for the original and also for

the slightly changed scenario vector. Contrary to [7], the scenario matrix X is assumed to be fixed. Let $\bar{\mathbf{y}} = (\bar{y}^1, ..., \bar{y}^T)$ be a perturbed scenario vector of returns of the tested asset. We consider the following distance between the original scenario vector and perturbed one:

$$d(\mathbf{y}, \bar{\mathbf{y}}) = \max_{1 \le t \le T} (\bar{y}^t - y^t)$$

.

For a given $\epsilon > 0$, we say that a tested asset is ϵ-SSD efficient if it is classified as SSD efficient for original scenario vector $\mathbf{y}$ as well as for all perturbed scenario vectors $\bar{\mathbf{y}}$ satisfying $d(\mathbf{y}, \bar{\mathbf{y}}) \le \epsilon$. Similarly, a tested portfolio is called ϵ-SSD inefficient if it is classified as SSD inefficient in the original case as well as in the case of perturbed scenarios that are from the ϵ-neighbourhood of $\mathbf{y}$.

Modifying Theorem 1, we can easily formulate a necessary and sufficient condition for ϵ-SSD efficiency of a given asset.

Theorem 2. *Let $z^t = y^t - \epsilon$, $t = 1, ..., T$ and*

$$\theta_E^* = \min_{\theta, \beta_t} \theta \tag{3}$$

$$\begin{aligned} s.t. \quad \sum_{t=1}^{T} \beta_t (z^t - x_n^t) + T\theta &\ge 0 \quad n = 1, 2, ..., N \\ \beta_t - \beta_{t+1} &\ge 0 \quad t = 1, 2, ..., T-1 \\ \beta_t &\ge 0 \quad t = 1, 2, ..., T-1 \\ \beta_T &= 1 \end{aligned}$$

A given asset $\boldsymbol{\tau}$ is ϵ-SSD efficient if and only if $\theta_E^* \le 0$.

Proof. The choice $z^t = y^t - \epsilon$, $t = 1, ..., T$ represents the worst case of returns from ϵ-neighbourhood of $\mathbf{y}$. Therefore, if $\theta_E^* \le 0$ for $z^t = y^t - \epsilon$, $t = 1, ..., T$ then $\theta_E^* \le 0$ for any $\bar{\mathbf{y}}$ from ϵ-neighbourhood of $\mathbf{y}$ and hence the tested asset is ϵ-SSD efficient. On the other hand, if $\theta_E^* > 0$ for $z^t = y^t - \epsilon$, $t = 1, ..., T$ then $\bar{\mathbf{y}} = \mathbf{y} - \epsilon\mathbf{1}$ is the perturbation causing that the tested asset is not ϵ-SSD efficient. □

The robust test from Theorem 2 is again based on solving linear program. Hence, it is easy to solve using any linear programming solver (algorithm).

If the tested asset is SSD inefficient, then it is not ϵ-SSD efficient for any $\epsilon > 0$. The test of ϵ-SSD inefficiency can be easily derived from the Post test as follows.

Theorem 3. *Let $z^t = y^t + \epsilon$, $t = 1, ..., T$ and*

$$\theta_I^* = \min_{\theta, \beta_t} \theta \tag{4}$$

$$\begin{aligned} s.t. \quad \sum_{t=1}^{T} \beta_t (z^t - x_n^t) + T\theta &\ge 0 \quad n = 1, 2, ..., N \\ \beta_t - \beta_{t+1} &\ge 0 \quad t = 1, 2, ..., T-1 \\ \beta_t &\ge 0 \quad t = 1, 2, ..., T-1 \\ \beta_T &= 1 \end{aligned}$$

A given asset $\boldsymbol{\tau}$ is ϵ-SSD inefficient if and only if $\theta_I^* > 0$.

The proof of Theorem 3 is very similar to the proof of Theorem 2, where $z^t = y^t + \epsilon$, $t = 1, ..., T$ is the best choice from ϵ-neighbourhood of $\mathbf{y}$.

4 Empirical study

To present the SSD efficiency tests, we apply them to historical US stock market data. We consider monthly excess returns from January 1982 to December 2011 ($T = 360$ observations) of $N = 48$ representative industry stock portfolios that serve as the base assets. The values are considered in percentage

representation. The industry portfolios are based on four-digit SIC codes and they are from Kenneth French data library. We test whether a US market portfolio, proxied by CRSP all share index is SSD and ϵ-SSD (in)efficient relative to all portfolios that can be created from the considered 48 representative US industry stock portfolios. Moreover, if the tie in scenarios of market portfolio returns occurs we slightly modify the tied scenarios (adding very small value) to have no ties. It is needed mainly in bootstrap application. This adjustments have no impact on the efficiency testing results.

Firstly, we apply Theorem 1 to test SSD portfolio efficiency of the market portfolio. We find $\theta^* = 1.0432$ and hence the market portfolio is classified as SSD inefficient. The value of θ^* shows that the minimal value (over all concave utility functions) of maximal violation (over all 48 base assets) of efficiency criteria is equal to 1.0432%. Therefore, the maximal ϵ for which the market portfolio is ϵ-SSD inefficient is $\epsilon = 1.0432$. We can easily check it also using Theorem 3. It means that if the returns of market portfolio are proxied with accuracy smaller or equal to 1.0432, that is, the differences between theoretical values and proxied values of all return's scenarios are smaller or equal to 1.0432, then the market portfolio is always classified as SSD inefficient.

Finally we construct 10 000 pseudo-samples of the same length as the original data and we do the bootstrap testing with replacement. More details about various resampling techniques can be found in [3]. We find that the market portfolio is very strongly SSD inefficient. The inefficiency is detected in all 10 000 pseudo-samples and the estimated (bootstrap) p-value of SSD efficiency of US market portfolio is smaller than 0.0001.

5 Conclusions

This paper deals with robustness and bootstrap techniques in SSD portfolio efficiency testing. It presents new ϵ-SSD portfolio (in)efficiency tests as a robust generalization of the Post test. These tests are very computationally attractive requiring only solving linear programming problems.

In empirical application, we analyze SSD efficiency of US stock market portfolio relative to all portfolios that can be created from 48 representative industry portfolios. We find that the market portfolio is SSD inefficient. Moreover, this classification remains the same if we slightly perturb the returns of market portfolio. To reach SSD efficiency of market portfolio, one has to increase each its scenario by at least 1.0432. And bootstrap techniques show that this minimal required increase is very high, because the market portfolio is classified as SSD inefficient in all 10 000 pseudo-samples.

Acknowledgements

The research was partially supported by the Czech Science Foundation under grants P402/10/1610 and P402/12/0558.

References

[1] Branda, M., Kopa, M. (2012): DEA-risk Efficiency and Stochastic Dominance Efficiency of Stock Indices. *Czech Journal of Economics and Finance*, 62, No. 2, 106–124.

[2] Branda, M., Dupačová, J. (2012): Approximations and contamination bounds for probabilistic programs. *Annals of Operations Research*, 193, No. 1, 3–19.

[3] Chernick, M.R. (2008): *Bootstrap Methods: A Guide for Practitioners and Researchers (Wiley Series in Probability and Statistics)*, John Wiley & Sons, Inc., New Jersey.

[4] Dupačová, J., Kopa, M. (2012): Robustness in stochastic programs with risk constraints. *Annals of Operations Research*, published on-line January 4, 2011, DOI:10.1007/s10479-010-0824-9.

[5] Fishburn, P.C. (1974): Convex stochastic dominance with continuous distribution functions. *Journal of Economic Theory* 7, 143–158.

[6] Hanoch, G., Levy, H. (1969): The efficient analysis of choices involving risk. *Review of Economic Studies*, 36, No. 3, 335–346.

[7] Kopa, M. (2010): Measuring of second-order stochastic dominance portfolio efficiency, *Kybernetika* 46, No. 3, 488–500.

[8] Kopa, M., Chovanec, P. (2008): A second-order stochastic dominance portfolio efficiency measure. *Kybernetika* 44, No. 2, 243–258.

[9] Kopa, M., Post, T. (2009): A portfolio efficiency test based on FSD optimality. *Journal of Financial and Quantitative Analysis* 44, No. 5, 1103–1124.

[10] Kuosmannen, T. (2004): Efficient diversification according to stochastic dominance criteria. *Management Science* 50, No. 10, 1390–1406.

[11] Levy, H. (2006): *Stochastic dominance: Investment decision making under uncertainty*. Second edition, Springer, New York.

[12] Markowitz, H.M. (1952): Portfolio selection. *The Journal of Finance* 7, No. 1, 77–91.

[13] Markowitz, H.M. (1959). *Portfolio Selection: Efficient Diversification in Investments*. John Wiley & Sons, Inc., New York.

[14] von Neumann, J., Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.

[15] Pflug, G. (2000). Some Remarks on the value-at-risk and the Conditional value-at-risk. In: *Probabilistic Constrained Optimization* (S.P. Uryasev, ed.), Kluwer, Dordrecht, 272–281.

[16] Post, T. (2003). Empirical tests for stochastic dominance efficiency. *Journal of Finance* 58, 1905–1932.

[17] Rockafellar, R.T., Uryasev, S. (2000). *Optimization of Conditional Value-at-Risk*. Journal of Risk, 2, 21–41.

[18] Rockafellar, R.T., Uryasev, S. (2002). *Conditional Value-at-Risk for General Loss Distributions*. Journal of Banking and Finance, 26, 1443-1471.

Characterization of uniformly quasi-concave functions

Miloš Kopa[1], Petr Lachout[2]

Abstract. Quasi-concave functions appears in economics and finance as utility functions, measures of risk or other objects, mainly in portfolio selection analysis. A special attention was paid to these functions in the minimax theory. Unfortunately, their limited application is due to the fact that supremum, sum, product of quasi-concave functions are typically not quasi-concave. This difficulty is removed by establishing of uniformly quasi-concave functions, due to Prékopa, Yoda and Subasi (2011). Supremum and sum of uniformly quasi-concave functions are also a quasi-concave function. Moreover, product of nonnegative uniformly quasi-concave functions is a quasi-concave function. We contribute with a new characterization of uniformly quasi-concave functions that allows for easier verification and provide more straightforward insight. Hence, application and usage of uniformly quasi-concave functions become to be easier and more natural.

Keywords: Quasi-concave function, uniformly quasi-concave functions, partial ordering, total ordering, monotonicity.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Let start with a definition of quasi-concave functions.

Definition 1. Let $E \subset \mathbb{R}^n$. We say that $f : E \to \mathbb{R}$ is quasi-concave if

1. E is convex.
2. For each $\alpha \in \mathbb{R}$ the level set $\mathsf{lev}_{\geq\alpha} f = \{x \in E \,:\, f(x) \geq \alpha\}$ is convex.

Alternatively, one can deal with quasi-convex functions. Function f is quasi-convex if and only if function $-f$ is quasi-concave. All these functions are useful in economics and finance, they serve as utility functions, measures of risk or other objects, mainly in portfolio selection analysis. In this paper we focus on quasi-concave functions. Unfortunately, their limited application is due to the fact that supremum, sum, product of quasi-concave functions are typically not quasi-concave. This difficulty is removed by establishing of uniformly quasi-concave functions, due to Prékopa, Yoda and Subasi (2011).

Definition 2. Let $E \subset \mathbb{R}^n$. Then, we say that functions $f_i : E \to \mathbb{R}$, $i = 1, 2, \ldots, m$ are

uniformly quasi-concave if

1. E is convex.
2. For each $i = 1, 2, \ldots, m$ the function f_i is quasi-concave.
3. For each $x, y \in E$ either

$$\forall i = 1, 2, \ldots, m \quad \min\{f_i(x), f_i(y)\} = f_i(x)$$

[1]Charles University in Prague/Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Praha 8, Czech Republic, e-mail: kopa@karlin.mff.cuni.cz
[2]Charles University in Prague/Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Praha 8, Czech Republic, e-mail: lachout@karlin.mff.cuni.cz

or

$$\forall i = 1, 2, \ldots, m \quad \min\{f_i(x), f_i(y)\} = f_i(y).$$

We present an equivalent descriptions of uniformly quasi-concave functions.

2 Equivalent characterizations of uniformly quasi-concave functions

Any set of functions determines a partial ordering on their common domain. This observation allows us to show announced characterizations.

Let us consider a subset $E \subset \mathbb{R}^n$ and a finite family of functions $\mathcal{F} = \{f_i, i = 1, 2, \ldots, m\} \subset \mathbb{R}^E$.

Definition 3. We say that functions of $\mathcal{F}$ are uniformly monotone if for each $x, y \in E$ either

$$\forall i = 1, 2, \ldots, m \quad \min\{f_i(x), f_i(y)\} = f_i(x)$$

or

$$\forall i = 1, 2, \ldots, m \quad \min\{f_i(x), f_i(y)\} = f_i(y).$$

We see that the functions of $\mathcal{F}$ are uniformly quasi-concave iff the functions of $\mathcal{F}$ are uniformly monotone and each of them is quasi-concave.

Definition 4. The set of functions $\mathcal{F}$ determines a partial ordering $\prec^{\mathcal{F}}$ and equivalence $\sim^{\mathcal{F}}$ on E by

$$x \prec^{\mathcal{F}} y \iff \begin{array}{l} \forall i = 1, 2, \ldots, m : f_i(x) \le f_i(y) \\ \exists j \text{ s.t. } f_j(x) < f_j(y) \end{array}$$

$$x \sim^{\mathcal{F}} y \iff \forall i = 1, 2, \ldots, m : f_i(x) = f_i(y)$$

The partial ordering is giving an equivalent description of uniform monotonicity.

Theorem 1. *Functions of $\mathcal{F}$ are uniformly monotone iff the factor space $E/_{\sim^{\mathcal{F}}}$ is totally ordered by $\prec^{\mathcal{F}}/_{\sim^{\mathcal{F}}}$, i.e. for each couple $x, y \in E$ just one from the three following relations holds*

$$x \prec^{\mathcal{F}} y, \; x \sim^{\mathcal{F}} y, \; y \prec^{\mathcal{F}} x.$$

Proof. We will prove the equivalence.

1. Let functions of $\mathcal{F}$ are uniformly monotone.
   Fix $x, y \in E$, $x \not\sim^{\mathcal{F}} y$.
   Then, there is an index j s.t. $f_j(x) \neq f_j(y)$.
   We have to distinguish two possibilities:
   (a) Let $f_j(x) < f_j(y)$.
   Hence from uniform monotonicity $\forall i = 1, 2, \ldots, m : f_i(x) \le f_i(y)$.
   Consequently, $x \prec^{\mathcal{F}} y$.
   (b) Let $f_j(x) > f_j(y)$.
   Hence from uniform monotonicity $\forall i = 1, 2, \ldots, m : f_i(x) \ge f_i(y)$.
   Consequently, $y \prec^{\mathcal{F}} x$.
   We have proved that the factor space $E/_{\sim^{\mathcal{F}}}$ is totally ordered by $\prec^{\mathcal{F}}/_{\sim^{\mathcal{F}}}$.
2. Let the factor space $E/_{\sim^{\mathcal{F}}}$ is totally ordered by $\prec^{\mathcal{F}}/_{\sim^{\mathcal{F}}}$.
   Fix $x, y \in E$.
   Since the factor space $E/_{\sim^{\mathcal{F}}}$ is totally ordered by $\prec^{\mathcal{F}}/_{\sim^{\mathcal{F}}}$, we have to distinguish three possibilities:

(a) If $x \sim^{\mathcal{F}} y$ then, $\forall i = 1, 2, \ldots, m$ $\min\{f_i(x), f_i(y)\} = f_i(x) = f_i(y)$.

(b) If $x \prec^{\mathcal{F}} y$ then, $\forall i = 1, 2, \ldots, m$ $\min\{f_i(x), f_i(y)\} = f_i(x)$.

(c) If $y \prec^{\mathcal{F}} x$ then, $\forall i = 1, 2, \ldots, m$ $\min\{f_i(x), f_i(y)\} = f_i(y)$.

We have shown that functions of $\mathcal{F}$ are uniformly monotone.

□

Characterization by partial ordering implies characterization by composition of appropriate functions.

Theorem 2. *The following statements are equivalent:*

1. *Functions of $\mathcal{F}$ are uniformly monotone.*
2. *There are a function $\psi : E \to \mathbb{R}$ and non-decreasing functions $\varphi_i : \psi(E) \to \mathbb{R}$, $i = 1, 2, \ldots, m$ such that $f_i = \varphi_i \circ \psi$ for all $i = 1, 2, \ldots, m$.*
3. *There are a function $\psi : E \to \mathbb{R}$ and non-decreasing functions $\varphi_i : \mathbb{R} \to \mathbb{R}$, $i = 1, 2, \ldots, m$ such that $f_i = \varphi_i \circ \psi$ for all $i = 1, 2, \ldots, m$.*

Proof. To prove announced equivalences, we prove step by step chain of implications.

1. Evidently, (3) ⇒ (2) and (2) ⇒ (1).
2. Let functions of $\mathcal{F}$ are uniformly monotone.

   We set $\psi = f_1 + f_2 + \cdots + f_m$.

   Hence,

$$\begin{aligned} x \prec^{\mathcal{F}} y &\iff \psi(x) < \psi(y), \\ x \sim^{\mathcal{F}} y &\iff \psi(x) = \psi(y), \\ y \prec^{\mathcal{F}} x &\iff \psi(x) > \psi(y). \end{aligned}$$

   For $i = 1, 2, \ldots, m$: we define $\varphi_i : \psi(E) \to \mathbb{R}$ such that for $d \in \psi(E)$ we set

$$\varphi_i(d) = f_i(x) \iff \psi(x) = d.$$

   The definition is correct because of

$$\psi(y) = d \Longleftrightarrow x \sim^{\mathcal{F}} y \Longleftrightarrow \forall i = 1, 2, \ldots, m : \ f_i(x) = f_i(y).$$

   (a) We have constructed a function $\psi : E \to \mathbb{R}$ and non-decreasing functions $\varphi_i : \psi(E) \to \mathbb{R}$, $i = 1, 2, \ldots, m$ such that $f_i = \varphi_i \circ \psi$ for all $i = 1, 2, \ldots, m$. Thus, statement (2) is fulfilled.

   (b) For each $i = 1, 2, \ldots, m$, we extend the function φ_i to the whole $\mathbb{R}$ by

$$\begin{aligned} \xi_i(t) &= \sup\{\varphi_i(d) \,:\, d \le t,\ d \in \psi(E)\} \quad \text{if} \quad \exists d \in \psi(E) \text{ s.t. } d \le t, \\ &= \inf\{\varphi_i(d) \,:\, d \in \psi(E)\} \quad \text{if} \quad \forall d \in \psi(E) : \ d > t. \end{aligned}$$

   The function ξ_i is non-decreasing and for all $d \in \psi(E)$ we have $\xi_i(d) = \varphi_i(d)$.

   The final tasks is to show that ξ_i does not reach infinite values.

   i. Assume $t \in \mathbb{R}$ such that $\xi_i(t) = -\infty$.
   Then, there is a sequence $d_n \in \psi(E)$, $n \in \mathbb{N}$ such that $d_n > t$ for all $n \in \mathbb{N}$ and $\varphi_i(d_n)$ is tending to $-\infty$.
   Consequently for any $x_n \in E$ such that $\psi(x_n) = d_n$, we observe $f_i(x_n) = \varphi_i(d_n)$ is tending to $-\infty$.
   According to definition of ψ and uniform monotonicity
   $\psi(x_n) = f_1(x_n) + f_1(x_n) + \cdots + f_1(m) = d_n$ is tending to $-\infty$.
   That is a contradiction since we assume $t \in \mathbb{R}$ with $d_n > t$ for all $n \in \mathbb{N}$.

ii. Assume $t \in \mathbb{R}$ such that $\xi_i(t) = +\infty$.
Then, there is a sequence $d_n \in \psi(E)$, $n \in \mathbb{N}$ such that $d_n \leq t$ for all $n \in \mathbb{N}$ and $\varphi_i(d_n)$ is tending to $+\infty$.
Consequently for any $x_n \in E$ such that $\psi(x_n) = d_n$, we observe $f_i(x_n) = \varphi_i(d_n)$ is tending to $+\infty$.
According to definition of ψ and uniform monotonicity
$\psi(x_n) = f_1(x_n) + f_1(x_n) + \cdots + f_1(m) = d_n$ is tending to $+\infty$.
That is a contradiction since we assume $t \in \mathbb{R}$ with $d_n \leq t$ for all $n \in \mathbb{N}$.

We have constructed a function $\psi : E \to \mathbb{R}$ and non-decreasing functions $\xi_i : \mathbb{R} \to \mathbb{R}$, $i = 1, 2, \ldots, m$ such that $f_i = \xi_i \circ \psi$ for all $i = 1, 2, \ldots, m$. Thus, statement (3) is fulfilled.

□

The observations can be summarized to give equivalent description for uniformly quasi-concave functions.

Theorem 3. *The following statements are equivalent:*

1. *Functions of $\mathcal{F}$ are uniformly quasi-concave.*
2. *Functions of $\mathcal{F}$ are uniformly monotone and each of them is quasi-concave.*
3. *Each function of $\mathcal{F}$ is quasi-concave and the factor space $E/_{\sim^{\mathcal{F}}}$ is totally ordered by $\prec^{\mathcal{F}}/_{\sim^{\mathcal{F}}}$.*
4. *There are a quasi-concave function $\psi : E \to \mathbb{R}$ and non-decreasing functions $\varphi_i : \psi(E) \to \mathbb{R}$, $i = 1, 2, \ldots, m$ such that $f_i = \varphi_i \circ \psi$ for all $i = 1, 2, \ldots, m$.*
5. *There are a quasi-concave function $\psi : E \to \mathbb{R}$ and non-decreasing functions $\varphi_i : \mathbb{R} \to \mathbb{R}$, $i = 1, 2, \ldots, m$ such that $f_i = \varphi_i \circ \psi$ for all $i = 1, 2, \ldots, m$.*

Proof. The proof is based on the observation that the functions of $\mathcal{F}$ are uniformly quasi-concave iff the functions of $\mathcal{F}$ are uniformly monotone and each of them is quasi-concave.

The other equivalent statements follow descriptions of uniform monotonicity derived in Theorem 1 and Theorem 2. Moreover, construction $\psi = f_1 + f_2 + \cdots + f_m$ in the proof of Theorem 2 is giving quasi-concave function ψ since sum of uniformly quasi-concave functions is quasi-concave, for proof see Prékopa, Yoda and Subasi (2011). □

3 Empirical examples

This section presents three examples of uniformly quasi-concave functions from decision making theory. These examples illustrate new characterization of quasi-concave functions in a simple way, where functions φ_i and ψ are easily identified. However, in general, the analytic prescription of φ_i and ψ may be much more demanding to find.

Example 1
Utility functions, introduced in von Neumann and Morgenstern (1944), are one of the basic tools of decision making theory. Especially, if a portfolio selection model maximizes the expected utility of the final wealth a proper choice of the particular utility function is very important. Depending on the investor's risk attitude we basically distinguish between three classes of utility functions: concave (suitable for risk averse investor), linear (risk neutral investor) and convex (risk seeking investor). In this example, we choose one utility function from each class, that is, $\mathcal{F} = \{u_1, u_2, u_3\}$ where $u_1(W) = e^W, u_2(W) = W, u_3(W) = -e^{-W}, \quad W \in \mathbb{R}$

Since all three utility functions are increasing, the uniform monotonicity of $\mathcal{F}$ can be easily verified. Moreover, each function of $\mathcal{F}$ is quasi-concave, because u_1 is an increasing convex function and u_2, u_3 are concave functions. Therefore we can apply Theorem 3 and find a simple characterization with quasi-concave inner function $\psi(W) = W$ and non-decreasing outer functions $\varphi_i(W) = u_i(W)$, $i = 1, 2, 3$.

Example 2
Following Pratt (1964) we may express the investor's risk attitude by (Arrow-Pratt) absolute risk aversion

measure (function) that is derived from twice differentiable utility function as the ratio of the second and the first derivative of the utility function:

$$r(W) = -\frac{u''(W)}{u'(W)}.$$

Since almost all investors are risk averse with decreasing absolute risk aversion measure we limit our attention on utility functions having Hyperbolic Absolute Risk Aversion (HARA functions):

$$r(W) = \frac{1}{aW+b}, \quad W \in \mathbb{R}^+$$

with $a, b > 0$. We add also the limiting cases when either $a = 0$ or $b = 0$. Summarizing, we again consider a family of three quasi-concave functions: $\mathcal{F} = \{r_1, r_2, r_3\}$ where

$$\begin{aligned} r_1(W) &= \frac{1}{aW+b}, \quad W \in \mathbb{R}^+, \quad a, b > 0 \\ r_2(W) &= \frac{1}{aW}, \quad W \in \mathbb{R}^+, \quad a > 0 \\ r_3(W) &= \frac{1}{b}, \quad W \in \mathbb{R}^+, \quad b > 0. \end{aligned}$$

The positive constants a, b guarantees uniform monotonicity of $\mathcal{F}$ because all functions in $\mathcal{F}$ are non-increasing. Moreover, all these functions are monotone and convex and hence quasi-concave. Summarizing, functions in $\mathcal{F}$ are uniformly quasi-concave. Using Theorem 3 we can represent them as follows $r_i = \varphi_i \circ \psi$, $i = 1, 2, 3$ where:

$$\begin{aligned} \psi(W) &= \frac{1}{W}, \quad W \in \mathbb{R}^+ \\ \varphi_1(x) &= \frac{x}{bx+a}, \quad x \in \mathbb{R}^+, a, b > 0 \\ \varphi_2(x) &= \frac{x}{a}, \quad x \in \mathbb{R}^+, a > 0 \\ \varphi_3(x) &= \frac{1}{b}, \quad b > 0. \end{aligned}$$

Of course the choice of functions φ_i, $i = 1, 2, 3$ and ψ is generally not unique and one can alternatively consider the following characterization:

$$\begin{aligned} \psi(W) &= \frac{b}{aW}, \quad W \in \mathbb{R}^+, a, b > 0 \\ \varphi_1(x) &= \frac{x}{xb+b}, \quad x \in \mathbb{R}^+, b > 0 \\ \varphi_2(x) &= \frac{x}{b}, \quad x \in \mathbb{R}^+, b > 0 \\ \varphi_3(x) &= \frac{1}{b}, \quad b > 0. \end{aligned}$$

In both cases the inner functions are quasi-concave and the outer functions are non-decreasing. Moreover, in the second case, the outer functions do not depend on a.

Example 3

Finally, we present the example of uniformly quasi-convex deviation measures. Deviation measures are derived from risk measures, see Rockafellar et al. (2006), that typically appear either in mean-risk portfolio selection models or in portfolio efficiency testing with respect to mean-risk criteria or stochastic dominance relation. See e.g. Branda and Kopa (2012), Kopa (2010) and references there in for more details. We consider normally distributed returns of M assets $\mathbf{r} \sim N(\mathbf{m}, V)$. Investor may combine these assets into portfolios with weights $\mathbf{w} \in \Lambda$. We do not allow short positions and impose the budget constraint, that is,

$$\Lambda = \{\mathbf{w} \in \mathbb{R}^M | \mathbf{1}'\mathbf{w} = 1, \quad w_i \geq 0, \quad i = 1, 2, ..., M\}.$$

We consider conditional Value at Risk at level $\alpha \in [0, 1]$ ($CVaR_\alpha$), introduced in Rockafellar and Uryasev (2000, 2002), as the suitable risk measure. For portfolio $\mathbf{w}$, it can be computed as follows:

$$CVaR_\alpha(Z) = \min_{\eta \in \mathbb{R}} \left\{ \eta + \frac{1}{1-\alpha} \mathbb{E}[Z - \eta]^+ \right\},$$

where $[\cdot]^+ = \max\{\cdot, 0\}$ denotes the positive part, $Z = -\mathbf{r}'\mathbf{w}$ and η is a real auxiliary variable. Since we assume normally distributed returns the formulation of conditional Value at Risk of portfolio $\mathbf{w}$ can be simplified to:

$$CVaR_\alpha(-\mathbf{r}'\mathbf{w}) = -\mathbf{m}'\mathbf{w} + g(\alpha)\sqrt{\mathbf{w}'V\mathbf{w}}$$

where $g : [0, 1] \to \mathbb{R}^+$ is increasing function, see Rockafellar and Uryasev (2000). Following Rockafellar et al. (2006) we construct the corresponding deviation measure ($DCVaR_\alpha$) as $CVaR$ of deviation between portfolio random loss and it expected value, that is:

$$DCVaR_\alpha(-\mathbf{r}'\mathbf{w}) = CVaR_\alpha(-\mathbf{r}'\mathbf{w} + \mathbf{m}'\mathbf{w}) = g(\alpha)\sqrt{\mathbf{w}'V\mathbf{w}}$$

The last equation follows from coherency of CVaR, see Artzner et. al (1999). As it is typical in risk shaping with CVaR (see Rockafellar and Uryasev (2002)) or stochastic dominance constraints (see e.g. Dentcheva and Ruszczynski (2006) or Kopa and Chovanec (2008)), we consider DCVaRs at particular levels α_k, $k = 1, 2, ..., K$. Each $DCVaR_{\alpha_k}(-\mathbf{r}'\mathbf{w})$ is quasi-convex because the variance matrix is positive-semidefinite. Since deviation measures are typically minimized and theory of quasi-concave function is formulated for maximization problems, we simply consider $f_k(\mathbf{w}) = -DCVaR_{\alpha_k} = -g(\alpha_k)\sqrt{\mathbf{w}'V\mathbf{w}}$, $k = 1, 2, ..., K$ that are quasi-concave functions. Moreover $g(\alpha_k)$ are always positive, and hence $\mathcal{F} = \{f_1(\mathbf{w}), ..., f_K(\mathbf{w})\}$ is uniformly monotone. Summarizing, $\mathcal{F}$ is uniformly quasi-concave and applying decomposition from Theorem 3, we can easily find quasi-concave inner function $\psi(\mathbf{w})$ and non-decreasing outer functions $\varphi_k(x)$:

$$\begin{aligned} \psi(\mathbf{w}) &= -\sqrt{\mathbf{w}'V\mathbf{w}}, \quad \mathbf{w} \in \Lambda \\ \varphi_k(x) &= g(\alpha_k)x, \quad x \in \mathbb{R}^+, k = 1, 2, ..., K. \end{aligned}$$

Acknowledgements

The research was supported by the Czech Science Foundation (grant P402/12/G097).

References

[1] Artzner, P., Delbaen, F., Eber, J.-M., Heath, D.: Coherent measures of risk. *Mathematical Finance*, **9** (1999), 203–228.

[2] 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.

[3] Dentcheva, D., Ruszczyński, A.: Portfolio Optimization with Stochastic Dominance Constraints, *Journal of Banking and Finance* **30**, 2 (2006), 433–451.

[4] Kopa, M., Chovanec, P.: A second-order stochastic dominance portfolio efficiency measure. *Kybernetika* **44**, 2 (2008), 243–258.

[5] Kopa, M.: Measuring of second-order stochastic dominance portfolio efficiency, *Kybernetika* **46**, 3 (2010), 488–500.

[6] Prékopa, A., and Yoda, K., and Subasi, M.M.: Uniform Quasi-Concavity in Probabilistic Constrained Stochastic Programming. Operations Research Letters **39**,1 (2011), 188–192.

[7] von Neumann, J., Morgenstern, O.: *Theory of games and economic behavior*. Princeton University Press, 1944.

[8] J. W. Pratt: Risk aversion in the small and in the large, *Econometrica* **32** (1964), 122–136.

[9] Rockafellar, R.T., Uryasev, S., Zabarankin, M.: Optimality Conditions in Portfolio Analysis with General Deviation Measures. *Mathematical Programming*, **108**, 2 (2006), 515–540.

[10] Rockafellar, R.T., Uryasev, S.: Optimization of Conditional Value-at-Risk. *Journal of Risk*, **2** (2000), 21–41.

[11] Rockafellar, R.T., Uryasev, S.: Conditional Value-at-Risk for General Loss Distributions. *Journal of Banking and Finance*, **26** (2002), 1443–1471.

Multistage risk-averse asset allocation with transaction costs

Václav Kozmík[1]

Abstract. This paper deals with asset allocation problems formulated as multistage stochastic programming models. Dynamic models allow rebalancing the portfolio multiple times before the final investment horizon is reached. The CVaR risk measure is used for its favorable properties and time consistent model is developed. The risk-aversion coefficients are separate for every stage, which covers a wide range of investment strategies. For example one could start with an aggressive investment policy, but focus on reducing the risk as the final stage approaches. The stock prices are assumed to be interstage independent and to follow lognormal distribution. The Stochastic Dual Dynamic Programming algorithm is then applied to solve the presented models. An extensive numerical study based on the data from Prague Stock Exchange compares the results obtained from static two stage models with the results from dynamic models having multiple stages. Both cases, with or without the transaction costs, are considered. The computational part of this work is realized in C++ language and uses CPLEX to solve the linear programs.

Keywords: Stochastic multistage programming, stochastic dual dynamic programming algorithm, Monte Carlo sampling, risk averse optimization.

JEL classification: C63, D81, G11
AMS classification: 49M25

1 Introduction

Mean-risk models are well-known instruments in portfolio selection analysis. These two-criteria models both maximize the portfolio mean return as well as minimize the risk which is linked to the stock market trading. Basics of portfolio selection theory using variance and semivariance as a measure of risk were published in the article [7] and the book [8] by Harry Markowitz already in 1950s. Since that time, many improvements have been proposed and implemented. There are new ways of measuring risk, let us mention at least Conditional Value at Risk [11] (CVaR), which will be used in our model. Moreover, several different distributions, for instance log-normal distribution, have been applied in mean-risk models.

The latest research is focused on dynamic models which allow rebalancing the portfolio multiple times before the final investment horizon is reached, we refer to the book [4] for an extensive overview of stochastic modeling in finance. In this paper, we present one of these models, asset allocation problem with transaction costs formulated as multistage stochastic program, inspired by the article [12]. Wide range of different investment strategies can be covered by the model, because risk-aversion settings can be adjusted separately for each stage.

We solve the model using stochastic dual dynamic programming algorithm (SDDP) which originated in the work of Pereira and Pinto [9]. SDDP-style algorithms rely on the assumption of stage-wise independence to provide good performance for problems with multiple stages. In order to apply the SDDP algorithm, a discrete scenario tree with finite number of nodes in each stage has to be built. This can be done from historical market data, by sampling from some continuous-type distribution or by sampling using the moment matching method, presented in [5]. New problems arise in the multistage setting, such as how many descendants should be used for each stage, how to deal with the scenario count explosion or how to estimate the distributions for every stage. From many articles dealing with these and associated problems we refer at least to [2], [3] and [10].

[1]Charles University in Prague, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Prague 8, Czech Republic, e-mail: vkozmik@gmail.com

We organize the remainder of this article as follows: We present our risk-averse multi-stage model in the next section and provide a brief description of the SDDP algorithm in Section 3. Computational results are presented in Section 4 and we conclude and discuss ideas for future work in Section 5.

2 Asset-allocation model

Presented model is inspired by the articles [12] and [14]. Evolution of the assets price in stages $t = 2, \dots, T$ is supposed to be random, we denote $\boldsymbol{\xi}_t = \boldsymbol{p}_t$ as the ratio between the price of the assets in stage t and stage $t-1$. The first stage portfolio value is set to be equal to 1 and only the probability distributions governing future realizations $\boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_T$ are assumed known, independent of each other. The realization of $\boldsymbol{\xi}_2$ is known when decisions $\boldsymbol{x}_2$ must be made and so on up to the stage T. The decisions vectors $\boldsymbol{x}_t$ contain the optimal asset allocation strategy, while the term $\boldsymbol{p}_t \boldsymbol{x}_{t-1}$ captures the state of the system, meaning the total value of the assets in our portfolio. The history of the data process up to time t will be denoted $\boldsymbol{\xi}_{[t]}$, meaning $\boldsymbol{\xi}_{[t]} = (\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_t)$.

Our model gives user the possibility to choose risk aversion coefficient and confidence level $\lambda_t, \alpha_t \in (0, 1)$ separately for each stage of the model. In order to give a nested formulation of the model we will introduce following function which calculates expectation and risk coming from a random loss Z:

$$\rho_{t,\boldsymbol{\xi}_{[t-1]}}[Z] = (1-\lambda_t)\,\mathbb{E}\left[Z|\boldsymbol{\xi}_{[t-1]}\right] + \lambda_t\,\mathrm{CVaR}_{\alpha_t}\left[Z|\boldsymbol{\xi}_{[t-1]}\right]. \tag{1}$$

We will suppose that short-selling of assets is not allowed. The risk-averse multistage asset-allocation model with T stages can be written in the following nested form:

$$\begin{aligned}\min_{\mathbf{1}^T\boldsymbol{x}_1=1,\boldsymbol{x}_1\geq 0} \rho_{2,\boldsymbol{\xi}_{[1]}}\Bigg[\min_{\boldsymbol{p}_2^T\boldsymbol{x}_1=\mathbf{1}^T\boldsymbol{x}_2,\boldsymbol{x}_2\geq 0} -\mathbf{1}^T\boldsymbol{x}_2 + \rho_{3,\boldsymbol{\xi}_{[2]}}\Bigg[\min_{\boldsymbol{p}_3^T\boldsymbol{x}_2=\mathbf{1}^T\boldsymbol{x}_3,\boldsymbol{x}_3\geq 0} -\mathbf{1}^T\boldsymbol{x}_3 + \cdots \\ + \cdots + \rho_{T,\boldsymbol{\xi}_{[T-1]}}\left[\min_{\boldsymbol{p}_T^T\boldsymbol{x}_{T-1}=\mathbf{1}^T\boldsymbol{x}_T,\boldsymbol{x}_T\geq 0} -\mathbf{1}^T\boldsymbol{x}_T\right]\Bigg]\Bigg].\end{aligned} \tag{2}$$

Presented model differs from the other possible approaches by taking the risk measure as a function of the recourse value at each stage. It provides easy and explanatory policy construction rules and by nesting the optimal values we achieve the property of time consistency, which is also an important indicator about the model quality, see [15]. The interpretation of the objective function is not straightforward, but it can be viewed as the cost we would be willing to pay at the first stage instead of incurring the sequence of random costs $Z_1, \dots, Z_T$, see [13].

Our model also allows dynamic programming equations to be developed, for the detailed description we refer again to the article [14]. Using the definition of Conditional Value at Risk:

$$\mathrm{CVaR}_\alpha[Z] = \min_u u + \frac{1}{\alpha}\mathbb{E}[Z-u]_+ \tag{3}$$

in (1) the recourse value $Q_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$ at stage $t = 2, \dots, T$ equals the optimal value of the problem:

$$\begin{aligned}\min_{\boldsymbol{x}_t, u_t}\ & -\mathbf{1}^T\boldsymbol{x}_t + \lambda_{t+1}u_t + \mathcal{Q}_{t+1}(\boldsymbol{x}_t, u_t) \\ \text{s. t. } & \boldsymbol{p}_t^T\boldsymbol{x}_{t-1} = \mathbf{1}^T\boldsymbol{x}_t \\ & \boldsymbol{x}_t \geq 0,\end{aligned} \tag{4}$$

with $\mathcal{Q}_{T+1}(\cdot) \equiv 0$ and recourse function calculated from the following equation:

$$\mathcal{Q}_{t+1}(\boldsymbol{x}_t, u_t) = \mathbb{E}\left[(1-\lambda_{t+1})\,Q_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}) + \frac{\lambda_{t+1}}{\alpha_{t+1}}\left[Q_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}) - u_t\right]_+\right]. \tag{5}$$

Please note that in contrast with the risk-neutral multistage problem, additional decision variable u_t is introduced to represent the estimated Value at Risk level, see (3). The recourse function does not

depend on $\boldsymbol{\xi}_{[t]}$, because the process is assumed to be stage-wise independent. First stage optimal solution of the stochastic program is given by:

$$\begin{aligned} &\min_{\boldsymbol{x}_1, u_1} \ \lambda_2 u_1 + \mathcal{Q}_2(\boldsymbol{x}_1, u_1) \\ &\text{s. t. } \mathbf{1}^T \boldsymbol{x}_1 = 1 \\ &\qquad \boldsymbol{x}_1 \geq 0. \end{aligned} \tag{6}$$

After the introduction of auxiliary variables u_t, the problem was converted to a case that resembles the risk-neutral model. However, there is still nonlinear function present in our equations (4)–(5), and the standard version of the SDDP algorithm cannot be directly applied.

We still have to extend our model to cover possible transaction costs. We will consider the case where transaction costs are proportional to the value of the assets sold or bought, f_{tr} denoting the relative cost. The balancing equation between stage $t-1$ and stage t portfolios has to be modified to include the total cost of $f_{tr}\mathbf{1}^T|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}|$. Converting to linear formulation we replace the equations (4) with following dynamic programming equations ($t = 2, \ldots, T$) :

$$\begin{aligned} &\min_{\boldsymbol{x}_t, u_t} \ -\mathbf{1}^T \boldsymbol{x}_t + \lambda_{t+1} u_t + \mathcal{Q}_{t+1}(\boldsymbol{x}_t, u_t) \\ &\text{s. t. } \boldsymbol{p}_t^T \boldsymbol{x}_{t-1} - f_{tr}\mathbf{1}^T \boldsymbol{z}_t = \mathbf{1}^T \boldsymbol{x}_t \\ &\qquad \boldsymbol{z}_t \geq \boldsymbol{x}_t - \boldsymbol{x}_{t-1} \\ &\qquad \boldsymbol{z}_t \geq \boldsymbol{x}_{t-1} - \boldsymbol{x}_t \\ &\qquad \boldsymbol{x}_t \geq 0. \end{aligned} \tag{7}$$

We suppose that all the initial capital is used for investment into stocks, meaning transaction costs are constant for the first stage and can be therefore omitted. The first-stage stochastic program remains in the following form:

$$\begin{aligned} &\min_{\boldsymbol{x}_1, u_1} \ \lambda_2 u_1 + \mathcal{Q}_2(\boldsymbol{x}_1, u_1) \\ &\text{s. t. } \mathbf{1}^T \boldsymbol{x}_1 = 1 \\ &\qquad \boldsymbol{x}_1 \geq 0. \end{aligned} \tag{8}$$

3 Stochastic dual dynamic programming

We provide only a brief description of the algorithm and refer again to the articles [9] and [14] for more details. The algorithm performs series of iterations until satisfactory solution is found, meaning that some stopping rule, given lower and upper bound, is satisfied. We solve the following problem at stages $t = 2, \ldots, T$:

$$\begin{aligned} &\min_{\boldsymbol{x}_t, u_t} \ -\mathbf{1}^T \boldsymbol{x}_t + \lambda_{t+1} u_t + \theta_t \\ &\text{s. t. } \boldsymbol{p}_t^T \boldsymbol{x}_{t-1} - f_{tr}\mathbf{1}^T \boldsymbol{z}_t = \mathbf{1}^T \boldsymbol{x}_t \\ &\qquad \boldsymbol{z}_t \geq \boldsymbol{x}_t - \boldsymbol{x}_{t-1}, \ \boldsymbol{z}_t \geq \boldsymbol{x}_{t-1} - \boldsymbol{x}_t \\ &\qquad \theta_t \geq \hat{\mathcal{Q}}_t^j + \left(\mathbf{g}_t^j\right)^T (\boldsymbol{x}_t, \, u_t) \quad j = 1, \ldots, C \\ &\qquad \boldsymbol{x}_t \geq 0. \end{aligned} \tag{9}$$

Here, θ_t in the objective function, coupled with cut constraints $\theta_t \geq \hat{\mathcal{Q}}_t^j + \left(\mathbf{g}_t^j\right)^T (\boldsymbol{x}_t, \, u_t)$, forms the outer linearization of the recourse function $\mathcal{Q}_{t+1}(\boldsymbol{x}_t, u_t)$. The structural and nonnegativity constraints simply repeat the same constraints from the model. In the final stage T cut constraints and θ_t term in objective function are omitted.

During a typical iteration of the SDDP algorithm, cuts have been accumulated at each stage. On a forward pass we sample a number of linear paths through the tree. As we solve a sequence of problems along these forward paths, the cuts that have been accumulated so far are used to form decisions at each stage. The solutions found form a policy which does not anticipate the future. The costs incurred along

all the sampled forward paths through the tree can be used to estimate the expected cost of the current policy, thus providing the upper bound.

In the backward pass of the algorithm, we add cuts to the collection defining the current approximation of the future cost functions. We do this by solving the descendant nodes of each node in the linear paths from the forward pass, except in the final stage T. To form a cut, we use the objective values and subgradients of the descendant nodes and employ the chain rule to calculate the subgradient of function $\mathcal{Q}_{t+1}(\boldsymbol{x}_t, u_t)$. The cuts collected at any node in stage t apply to all the nodes in that stage, therefore only one set of cuts in maintained for each stage. This complexity reduction is possible because of the interstage independence assumption. The optimal value of the first-stage problem provides the lower bound.

4 Empirical study

We used weekly data of the most important assets traded on the Prague Stock Exchange, November 2007 to March 2012. The week-to-week ratios were adjusted to include the stock dividends. The data summary can be found in the Table 1. We have fitted multidimensional correlated log-normal distribution to the adjusted price ratios, and the scenario tree was then constructed by sampling from this distribution, using the polar method for normal distribution sampling. The L'Ecuyer random generator was used to generate the required uniform random variables. Every computation including the sampling process was repeated 10 times, allowing to compute standard deviations of the solutions and objective values. There are many software products available for the purpose of stochastic optimization, for an overview see [6] or [16], but there is nothing available for the risk-averse version of the SDDP algorithm. Our computation was therefore implemented in own C++ software, using CPLEX to solve the required linear programs and Armadillo library for matrix computations. The confidence level was always set to 5%.

asset	**mean**	**std. deviation**
AAA	0.9980	0.0716
CETV	0.9929	0.0995
ČEZ	0.9994	0.0406
ERSTE GROUP BANK	0.9983	0.0795
KOMERČNÍ BANKA	1.0018	0.0543
ORCO	0.9899	0.0938
PEGAS NONWOVENS	0.9995	0.0398
PHILIP MORRIS ČR	1.0035	0.0368
TELEFÓNICA C.R.	1.0004	0.0266
UNIPETROL	0.9986	0.0506

Table 1: Data summary

We evaluated the model with two different settings of risk coefficients, $\lambda_t = \frac{1}{2}$ and $\lambda_t = \frac{t-1}{T}$. The first represents neutral and stable risk-aversion, while the latter means that we want to be sure about the final portfolio value by being more risk-averse as the final stage approaches. Both cases, with transaction costs of 0.3% and without transaction costs were considered. We have computed the optimal first-stage decisions for models with 2, 3 and 5 stages. In the following Table 2 we show the setup for the scenario trees used in our algorithm.

stages	**descendants per node**	**total scenarios**
2	$50,000$	$50,000$
3	$1,000$	$1,000,000$
5	$1,000$	10^{12}

Table 2: Testing problems setup

In all the testing cases, only three assets play a significant role in our portfolio: ČEZ, PHILIP MORRIS ČR and TELEFÓNICA C.R.. We will exclude the other assets from our results to ease the

orientation. We present the results without any transaction costs first, showing the optimal first-stage decisions for constant and growing risk-aversion settings in Tables 3 and 4. The difference between the optimal portfolios for 2, 3 or 5 stages is not significant in the constant risk coefficients setting. In the second case we can see a slight movement to the riskier asset in the first-stage decision, putting more weight to PHILIP MORRIS ČR. This is expected as the risk settings target mostly the expectation part of our recourse function in the first stage.

stages	ČEZ	PHILL	TELE
2	0.0663 (0.0087)	0.3169 (0.0081)	0.6168 (0.0092)
3	0.0510 (0.0459)	0.3112 (0.0537)	0.6273 (0.0707)
5	0.0450 (0.0307)	0.3340 (0.0268)	0.6043 (0.0571)

Table 3: Optimal decisions (std. deviations) with $f_{tr} = 0$ and $\lambda_t = \frac{1}{2}$

stages	ČEZ	PHILL	TELE
2	0.0663 (0.0087)	0.3169 (0.0081)	0.6168 (0.0092)
3	0.0597 (0.0645)	0.3429 (0.0650)	0.5792 (0.0920)
5	0.0392 (0.0415)	0.4325 (0.0678)	0.4975 (0.0652)

Table 4: Optimal decisions (std. deviations) with $f_{tr} = 0$ and $\lambda_t = \frac{t-1}{T}$

Next we show the optimal first-stage decisions with transaction costs of 0.3% in the Tables 5 and 6. It should be noted that in accordance with our model transaction costs have no effect in 2-stage models. We observe that the presence of the transaction costs reduces the differences found in the previous case with varying risk coefficients. We believe that this follows from the fact that varying risk coefficients require the investor to change the portfolio in every stage significantly. However, with the transaction costs in mind, this could be more expensive than the loss coming from holding slightly suboptimal, but stable portfolio. The impact of the transaction costs should be weaker in cases where stages cover longer time periods instead of just weeks.

stages	ČEZ	PHILL	TELE
2	0.0663 (0.0087)	0.3169 (0.0081)	0.6168 (0.0092)
3	0.0405 (0.0279)	0.2977 (0.0322)	0.6438 (0.0409)
5	0.0643 (0.0208)	0.3115 (0.0231)	0.6149 (0.0323)

Table 5: Optimal decisions (std. deviations) with $f_{tr} = 0.3\%$ and $\lambda_t = \frac{1}{2}$

stages	ČEZ	PHILL	TELE
2	0.0663 (0.0087)	0.3169 (0.0081)	0.6168 (0.0092)
3	0.0412 (0.0389)	0.3175 (0.0258)	0.6192 (0.0403)
5	0.0493 (0.0240)	0.3274 (0.0346)	0.6168 (0.0293)

Table 6: Optimal decisions (std. deviations) with $f_{tr} = 0.3\%$ and $\lambda_t = \frac{t-1}{T}$

The impact of adding more stages to the stage-independent model with constant risk-aversion settings tends to be minimal. This could, however, be different in the stage-dependent case. On the other hand, varying risk coefficients provide distinct solutions even with the independence assumption. In order to handle stage-dependent models, we would need to use different algorithm to solve the models and more complex estimation procedures.

5 Conclusion

We have developed a simple stock asset allocation model for the multistage setting and successfully employed the SDDP algorithm to solve it. We provide results based on assets from Prague Stock Exchange

for 2, 3 and 5-stage problems. Our results show that the nested model with constant risk-aversion coefficients provides similar results in both the 2-stage and multistage setting. On the other hand, it provides distinct behavior in the case when varying risk coefficients are used to represent more complicated investment strategies. We have also evaluated the effect of transaction costs in our model and pointed out that their presence could draw the dynamic model behavior closer to the static one.

The future experiments should include problems with more stages as well as different time periods represented by the stages, especially months and years. We would also like to focus on recent ways of measuring risk, see for example [1]. The stage-dependent case should also be covered, although it requires the use of different algorithm than SDDP.

Acknowledgements

This research was supported by Czech Science Foundation (grant 402/12/G097).

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] Dupačová, J., Consigli, G. and Wallace, S. W.: Scenarios for multistage stochastic programs, *Annals of Operations Research* **100** (2000), 25–53.

[3] Dupačová, J., Gröwe-Kuska, N., Römisch, W.: Scenario reduction in stochastic programming: An approach using probability metrics, *Mathematical Programming* **95** (2003), 493–511.

[4] Dupačová, J., Hurt, J. and Štěpán, J.: *Stochastic Modeling in Economics and Finance*, Kluwer Academic Publishers, 2002.

[5] Hoyland, K., Kaut, M. and Wallace, S.: A Heuristic for Moment-Matching Scenario Generation, *Computational Optimization and Applications* **24 (2-3)** (2003), 169–185.

[6] Kopa, M., ed.: *On Selected Software for Stochastic Programming*, MatfyzPress, Prague, 2008.

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

[8] Markowitz, H. M.: *Portfolio Selection: Efficient Diversification of Investments*, John Wiley & Sons, New York, 1959.

[9] Pereira, M. V. F. and Pinto, L. M. V. G.: Multi-stage stochastic optimization applied to energy planning, *Mathematical Programming* **52** (1991), 359–375.

[10] Pflug, G. Ch.: Scenario tree generation for multiperiod financial optimization by optimal discretization, *Mathematical Programming* **89** (2001), 251–271.

[11] Rockafellar, R.T. and Uryasev, S.: Optimization of Conditional Value-At-Risk, *The Journal of Risk* **2 (3)** (2000), 21–41.

[12] Ruszczynski, A. and Shapiro, A.: Conditional Risk Mappings, *Mathematics of Operations Research* **31 (3)** (2006), 544–561.

[13] Ruszczynski, A.: Risk-averse dynamic programming for Markov decision processes, *Mathematical Programming* **125 (2)** (2010), 235–261.

[14] Shapiro, A.: Analysis of stochastic dual dynamic programming method, *European Journal of Operational Research* **209** (2011), 63–72.

[15] Shapiro, A.: On a time consistency concept in risk averse multistage stochastic programming, *Operations Research Letters* **37** (2009), 143–147.

[16] Wallace, S. W., Ziemba, W. T., eds.: *Applications of Stochastic Programming*, SIAM - Mathematical Programming Society Series on Optimization, 2005.

An agent-based model of price flexing by chain-store retailers

Ondřej Krčál [1]

Abstract. This paper investigates the effects of local price flexing on market structure and welfare in the supermarket sector. It presents an agent-based model of a sector with two chain-store retailers in which the number and location of stores and prices charged by the stores are determined endogenously. The outcome in which all the stores within each chain charge the same price is compared to the situation in which each store sets prices according to local market conditions. The paper finds that local pricing reduces the number of stores and total welfare in the market. Furthermore, local pricing is more likely to increase the average prices and total revenue in the market, if the reservation price of consumers is high relative to the equilibrium price under uniform pricing. In this situation, local pricing is likely to reduce not only total welfare but also aggregate consumer surplus in the market.

Keywords: retail, supermarket, price flexing, agent-based, local, uniform.

JEL classification: L10, L40, L81, R30
AMS classification: 37M05, 91B26

1 Introduction

Local price flexing (or local pricing) by chain-store retailers is a form of third-degree price discrimination in which individual stores set their prices according to their local market power. Prices are typically higher in areas with higher geographical distance between competing stores. In its investigation of the UK supermarket sector in 2000, Competition Commission found evidence of local price flexing. Moreover, it concluded that this practice distorted competition and adversely affected public interest. One of the remedies considered by Competition Commission was the imposition of uniform pricing, so that all stores within one chain would have to charge the same prices for the same products (for a detailed account of the use of local price flexing in retail markets, see [2], [3], and [4]).

This paper investigates the effects of local pricing on market outcome in the retail sector. This problem is closely related to the literature on third-degree price discrimination in oligopolistic markets ([1] and [5]). A more specific theoretical approach to the problem of local price flexing was proposed by Dobson & Waterson [2] and [3]. In [3], they present a stylized model of a supermarket sector with two retailers. Each of the retailers operates in two separate local markets: they compete against each other in one of the markets (there is a differentiated Bertrand competition) and have local monopoly in the other. Using a linear demand specification, they find that the results depend on two parameters: a measure of substitutability between the products sold in the common market and the relative size of the demand in the local and common market. They show that if the demand functions are similar in both markets, local pricing increases total industry profit for highly substitutable products and reduces it if the substitutability is low. They also show that for similar demand functions and high and intermediate subsitutability, local pricing reduces aggregate consumer surplus in the market.

One of the problems of the approach by Dobson & Waterson [2] and [3] is that pricing strategy affects only prices while leaving the market structure unchanged. This paper proposes a solution to this problem. It presents a version of the agent-based model of a monopolisticaly competitive market with endogenous number and location of firms introduced by Krčál [6]. In the model presented here, local pricing affects not only prices charged by the stores but also the number and location of stores in the market. Therefore, the model provides a more realistic framework for analyzing the effects of local pricing in retail sector. The structure of the paper is straightforward. Section 2 introduces the model, Section 3 describes the data and discusses the results of the model, and Section 4 concludes.

[1] Masaryk University/FEA, Economics department, Lipová 41a, 602 00 Brno, krcalo@mail.muni.cz

2 Model

In agent-based models, agents act or interact according to specific rules that are followed in a given order. In this section, I introduce the agents and explain the rules of the model in a step-by-step way similar to the order in which the simulation of the model proceeds (the model was implemented in the multiagent modeling environment `Netlogo 4.1.3`). In each run, the model is first initialized and then it runs for a certain number of periods.

In the initialization phase, the model creates a landscape, a population of consumers and an initial population of stores. *Landscape*: The retail market is set in a square landscape of 40×40 patches (a patch is a square field in Netlogo). In this model, a patch can be interpreted as any unit of distance. So, for instance, the landscape can be seen as a square island with a side of 40 kilometers. *Population of consumers*: The landscape is populated by 1,000 identical consumers who differ only in their locations. Consumers live in settlements. Each inhabitant of a settlement gets a location with a random direction from the center of the settlement and with a distance from the center of $\sqrt{h/(\pi u)}$, where h is the number of inhabitants of the settlement, $u > 0$ is the population-density parameter and $\pi \doteq 3.14$. If $u = 1$, the average population density of a round settlement with a perimeter of $\sqrt{h/(\pi u)}$ is 1 consumer per square patch. Only if a settlement is too close to the edge of the landscape, consumers who would be located outside of the landscape live on the edge of the square (or on the coastline of the square island), keeping their original direction from the center of the settlement. Then, the average population density in such a settlement is higher. *Initial population of stores*: The market is served by two chain-stores (chain 1 and chain 2) who sell identical product. Before the first period, ten stores of each chain locate randomly in the market. Each store sets its price equal to $p_R/2$, where $p_R > 0$ is the reservation price of consumers.

Once the model is initialized, the simulation proceeds in periods. In each period, the agents do their actions in four sequential steps: 1) Both chains consider opening new stores. 2) The existing stores adjust their prices. 3) Consumers shop in the stores. And 4) both chains consider closing some of their stores.

1) *Opening stores* At the beginning of period t, each chain considers building v new stores. They do it in turns. First, chain 1 and then chain 2 considers opening a store. This process is repeated v times. For each new store, each chain is assigned a random location. It would open the store in this location only if for given prices the new store increased the profit of the chain [the profit is determined in the same way as described in points 3) and 4)]. Whether a new store will enter the market therefore depends not only on its assigned location but also on the initial price it charges. The price depends on the pricing strategy of the chain. Under *uniform pricing* (strategy U), the new store charges the same price as any store in its chain. Under local pricing, the chain can assign any price to the new store. In order to test the sensitivity of the result to different pricing strategies of the new stores, I introduce three different local-pricing strategies: 1) Under *local pricing with minimal entry price* (strategy $\underline{L}$), the new store sets its price equal to the lowest price charged by an incumbent store of its chain. 2) Under *local pricing with average entry price* ($\hat{L}$), the new store sets its price equal to the average price charged by the stores of its chain. And 3) under *local pricing with local entry price* (L_L), the new store sets its price equal to the price the store (of any chain) with the lowest distance to the new store.

2) *Adjusting prices* After the new stores have been opened, each store can increase or decrease its price by a constant $\epsilon > 0$ or keep it at the same level as before. The adjustment process depends on the pricing strategy of the chain. Under uniform pricing (U), each store in a given chain charges the same price. Therefore, each *chain* chooses the price (out of the three options) that maximizes its profit given the price charged by the other chain. Under local pricing ($\underline{L}$, $\hat{L}$, or L_L), each store may charge a different price. Therefore, each *store* chooses the price that maximizes its chain's profit given the prices charged by all the other stores. Again the profit is determined in the way described in points 3) and 4).

3) *Shopping* Each consumer makes a decision whether to shop in store i based on the price of the product p_{it}, on the per-patch transportation cost $c > 0$ and on the distance to the store i d_{it}. Each consumer buys one unit of the product from the store with the lowest $p_{it} + cd_{it}^2$ if $p_R > p_{jt}^* + cd_{jt}^{*\,2}$, and zero units if $p_R \leq p_{jt}^* + cd_{jt}^{*\,2}$, where p_{jt}^* and d_{jt}^* are the price and the distance of the optimal store for consumer j in period t. The intuition behind the zero purchase is that each consumer can buy the product in a near-by mom-and-pop store for the price p_R. After the shopping and before the exit, all the variables used for analyzing the situation in the market are calculated and recorded.

4) *Closing stores* The closing decision is based on profit of stores. Assuming zero marginal cost, the profit of store i in period t is $\pi_{it} = q_{it}p_{it} - F$, where q_{it} are units of product sold, and F is the quasi-fixed cost that is the same for all the stores. In period t, the chain closes store i with a probability $-\pi_{it}/F$.

3 Results

This section shows how pricing strategy affects the situation in the retail market in the model. Subsection 3.1 provides details about the data and Subsection 3.2 presents the main results of the model. The data presented in this section was generated using the function Behavior Space in `Netlogo` and analyzed using the econometric software `Gretl 1.9.5cvs`.

3.1 Data

The data analyzed in this section come from the total number of 1,024 runs corresponding to all possible combinations of the following settings and parameter values: 2 types of landscape: 1) *urban landscape* containing 1 settlement with the center in the central point of the landscape and the number of inhabitants $h = 400$, and 20 settlements with randomly located centers and $h = 30$, and 2) *rural landscape* containing 1 settlement with the center in the central point of the landscape and $h = 100$, and 30 settlements with randomly located centers and $h = 30$; 2 population-density parameters $u = 0.5$ and $u = 1$; 2 reservation prices $p_R = 0.5$ and $p_R = 1$; 2 numbers of new stores per chain and period $v = 2$ and $v = 4$; 4 strategy profiles (U, U), $(\underline{L}, \underline{L})$, $(\hat{L}, \hat{L})$ and (L_L, L_L); 2 transportation-cost parameters $c = 0.01$ and $c = 0.02$; 2 price-change parameters $\epsilon = 0.02$ and $\epsilon = 0.03$; 1 quasi-fixed cost $F = 5$; and 4 random initializations with random seed of 1, 2, 3, and 4 (using the `random-seed` function in `Netlogo`).

Each run of the simulation generates the following variables:

- Quantity $Q = \frac{1}{100}\sum_{t=101}^{200} \bar{n}_t$, where $\bar{n}_t$ is the number of consumers who bought the product in one of the chains in period t (this consumers are called *customers*).
- Price $P = \frac{1}{100}\sum_{t=101}^{200}(\frac{1}{\bar{n}_t}\sum_{j=1}^{\bar{n}_t} p_{jt})$, where p_{jt} is the price paid by customer j in period t.
- Number of stores of chain k $M_k = \frac{1}{100}\sum_{t=101}^{200} m_{kt}$, where m_{kt} is the number of stores in chain k in period t.
- Revenue of chain k $R_k = \frac{1}{100}\sum_{t=101}^{200}\sum_{l=1}^{m_{kt}} q_{lkt}p_{lkt}$, where q_{lkt} and p_{lkt} are the quantity and price of store l of chain k in period t.
- Distance $D = \frac{1}{100}\sum_{t=101}^{200}\sum_{j=1}^{\bar{n}_t} d^*_{jt}$, where d^*_{jt} is the distance to the store of customer j in period t.
- Consumers' surplus $CS = \frac{1}{100}\sum_{t=101}^{200}\sum_{j=1}^{\bar{n}}(p_R - p_{jt} - cd^{*\,2}_{jt}) = Qp_R - R - cD^2$ where $R = R_1 + R_2$.
- Profit of chain k $\Pi_k = \frac{1}{100}\sum_{t=101}^{200}\sum_{l=1}^{m_{kt}}(p_{lkt}q_{lkt} - F) = R_k - M_kF$
- Total profit $\Pi = \Pi_1 + \Pi_2 = R - MF$, where $M = M_1 + M_2$.
- Welfare $W = CS + \Pi = Qp_R - cD^2 - MF$.

3.2 Comparing the market outcome for uniform and local pricing

This subsection compares the outcomes of the model with both chains pricing uniformly to the outcome of the model with both chains using local pricing. First, it explains how the outcomes will be compared. Then, it shows how local pricing affects pattern of prices and what are the effects of this change on quantity Q, number of stores M, and distance D. Finally, it investigates the effect of local pricing on welfare W, consumers' surplus CS and profit Π.

The market outcomes are compared as follows: There are three pairs of strategy profiles to be compared [(U, U) to $(\underline{L}, \underline{L})$, (U, U) to $(\hat{L}, \hat{L})$, and (U, U) to (L_L, L_L)]. For each of the pairs, I investigate the effect of local pricing on a given variables using linear OLS regressions. The dependent variables are regressed against all the exogenous variables of the model changed in the simulation (except for the random-seed variable). The independent variables are denoted as follows: transportation-cost parameter c is TRANSP_COST, population-density parameter u is POP_DENSITY, reservation price p_R is RES_PRICE, price-change parameter ϵ is EPSILON, the number of entrants per chain and period v is ENTRANTS, the variable that indicates the type of landscape is URBAN – it takes the value of 1 for urban landscape and 0 for rural landscape, and the variable that indicates pricing strategy is LOCAL – it takes the value of 1 for local pricing, and 0 for uniform pricing. For instance, the parameter of LOCAL in the following regression shows the effect of local pricing [strategy profile $(\underline{L}, \underline{L})$] on number of stores M:

$$\widehat{\text{STORES\_NO}} = \underset{(0.958)}{19.307} + \underset{(23.652)}{507.176}\,\text{TRANSP\_COST} - \underset{(0.473)}{3.454}\,\text{POP\_DENSITY} + \underset{(0.473)}{5.935}\,\text{RES\_PRICE}$$
$$+\underset{(23.652)}{19.293}\,\text{EPSILON} + \underset{(0.118)}{0.325}\,\text{ENTRANTS} - \underset{(0.237)}{1.336}\,\text{URBAN} - \underset{(0.237)}{4.523}\,\text{LOCAL} \qquad (1)$$
$$T = 512 \quad \bar{R}^2 = 0.677 \quad F(7, 504) = 153.64 \quad \hat{\sigma} = 2.676$$

(standard errors in parentheses)

For all other regressions the paper will report only the mean value of the dependent variable under uniform pricing, and the parameter and standard error of LOCAL. For instance, the paper will report 29.6 and −4.52 (0.24) for the equation 1 (see Table 1, column M). The meaning of the numbers is as follows: The change from uniform pricing (U, U) to local pricing with minimal entry price $(\underline{L}, \underline{L})$ reduces the mean value of number of stores M from 29.6 by 4.52 to roughly 25.1. The standard error is 0.24, which indicates a highly significant effect of pricing strategy on number of stores. Table 1 (section **all data**) summarizes the effect of local pricing on selected variables for the entire dataset. Each field in the rows $(\underline{L}, \underline{L})$, $(\hat{L}, \hat{L})$ and (L_L, L_L) corresponds to one regression. The total number of regressions run for the entire dataset is 24. Furthermore in order to test sensitivity of the results, I run the same 24 regressions for each of the 12 different partitions of the data (with 256 observations each) restricted to TRANSP_COST = 0.01 or 0.02, POP_DENSITY = 0.5 or 1, RES_PRICE = 0.5 or 1, EPSILON = 0.02 or 0.03, ENTRANTS = 2 or 4, and URBAN = 0 or 1. The parameter that affects most the outcome of the model is reservation price. For this reason, Table 1 (sections $\mathbf{p_R = 0.5}$ and $\mathbf{p_R = 1}$) also presents the effects of local pricing for the partitions restricted to RES_PRICE = 0.5 and 1. In most of the regressions, White test or Beusch-Pagan test indicate heteroskedasticity. All the regressions in this paper therefore report heteroskedasticity-robust standard errors.

Dataset	Pricing	Q	M	cD^2	W	P	R	CS	Π
all data ($T = 512$)	(U, U)	965.4	29.6	64.7	519.9	0.285	275.2	392.5	127.4
	$(\underline{L}, \underline{L})$	−27.8	−4.52	35.3	−29.1	−0.004	−10.1	−41.6	12.5
		(2.94)	(0.24)	(1.23)	(1.38)	(0.003)	(3.10)	(3.21)	(2.47)
	$(\hat{L}, \hat{L})$	−9.23	−1.59	20.2	−17.6	−0.013	−15.1	−10.6	−7.09
		(2.63)	(0.24)	(0.84)	(1.26)	(0.003)	(2.92)	(2.78)	(2.11)
	(L_L, L_L)	−10.8	−1.39	14.09	−13.2	0.002	−0.57	−19.6	6.38
		(2.69)	(0.24)	(0.76)	(1.21)	(0.003)	(3.12)	(2.89)	(2.24)
$\mathbf{p_R = 0.5}$ ($T = 256$)	(U, U)	932.0	28.1	56.7	269.0	0.262	243.6	165.8	103.2
	$(\underline{L}, \underline{L})$	−45.6	−4.55	17.8	−17.8	−0.028	−37.0	−3.51	−14.3
		(3.27)	(0.29)	(0.71)	(1.64)	(0.003)	(2.46)	(2.32)	(1.63)
	$(\hat{L}, \hat{L})$	−15.5	−2.12	12.2	−9.33	−0.023	−25.6	5.71	−15.1
		(2.77)	(0.26)	(0.57)	(1.63)	(0.003)	(2.36)	(2.38)	(1.67)
	(L_L, L_L)	−18.6	−2.26	7.63	−5.63	−0.013	−17.3	0.34	−5.98
		(2.79)	(0.27)	(0.57)	(1.58)	(0.003)	(2.54)	(2.53)	(1.69)
$\mathbf{p_R = 1}$ ($T = 256$)	(U, U)	998.8	31.0	306.7	72.8	770.8	0.307	619.3	151.5
	$(\underline{L}, \underline{L})$	−9.99	−4.49	52.9	−40.4	0.02	16.8	−79.7	39.3
		(0.71)	(0.32)	(1.31)	(1.99)	(0.004)	(3.81)	(4.05)	(2.98)
	$(\hat{L}, \hat{L})$	−3.01	−1.07	28.3	−26.0	−0.003	−4.47	−26.8	0.86
		(0.33)	(0.33)	(0.88)	(1.79)	(0.004)	(4.06)	(4.06)	(2.91)
	(L_L, L_L)	−2.88	−0.51	20.5	−20.8	0.017	16.2	−39.6	18.7
		(0.31)	(0.34)	(0.81)	(1.72)	(0.004)	(4.28)	(4.19)	(3.07)

Table 1 The effect of local pricing different dependent variables
The table shows the means of different dependent variables [see the lines denoted by (U, U)], and the parameters and standard errors (in parentheses) of the dummy variable LOCAL for three different local pricing strategies [see the lines denoted by $(\underline{L}, \underline{L})$, $(\hat{L}, \hat{L})$, and (L_L, L_L)]. This information is reported for the entire dataset (all data) and two different partitions of the dataset ($p_R = 0.5$ and $p_R = 1$).

Pricing strategy affects the prices charged by stores within each chain. While under uniform pricing all stores in a given chain charge the same price, the prices charged by stores under local pricing differ. The average difference between the highest and lowest price charged within a chain is approximately 0.24 for the reservation price $p_R = 0.5$, and 0.54 for the reservation price $p_R = 1$. Moreover, local pricing affects the geographical pattern of prices. The local level of prices depends on the density of population and other characteristics of the area. Typically, stores in larger and more densely populated areas charge lower prices. Figure 1 shows a typical spatial pattern of prices for $p_R = 0.5$ (left panel) and $p_R = 1$ (right panel) for the pricing strategy (L_L, L_L). Because of lower reservation price, lower proportion of customers in the left panel have prices higher than 0.3. And also, many of consumers living in relatively remote areas in this panel do not shop in the stores at all (see the dots in the left panel of Figure 1).

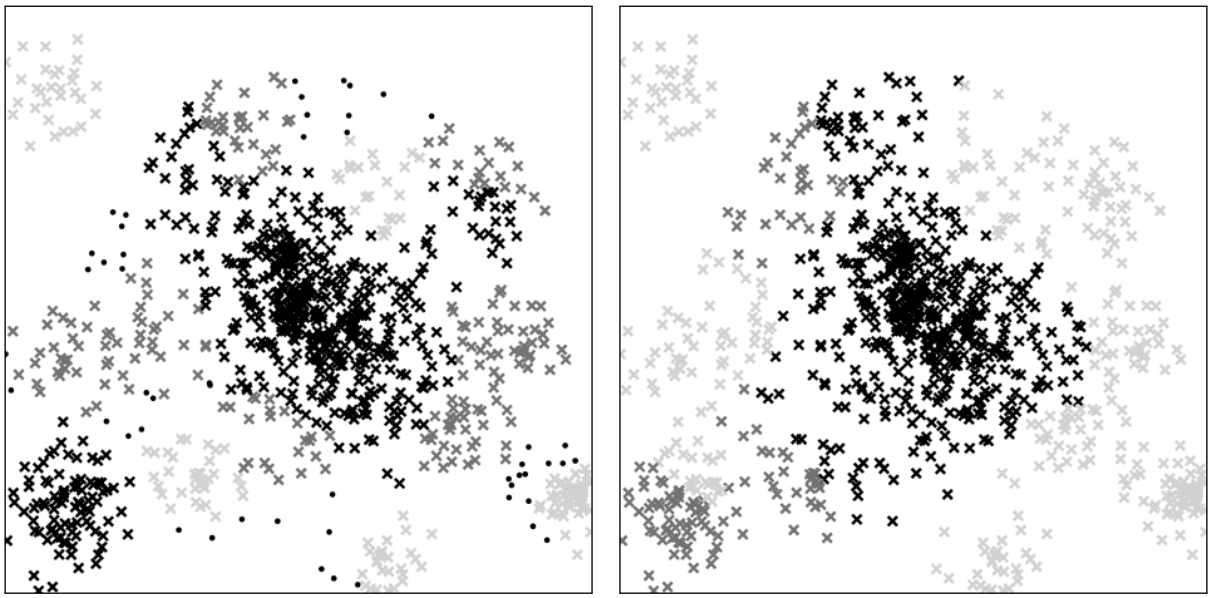


Figure 1 Examples of pattern of prices for $p_R = 0.5$ (left panel) and $p_R = 1$ and the strategy (L_L, L_L)
The panels show the market outcome in a market with the size of 40×40 patches in period $t = 200$. The crosses and dots are consumers located in the landscape. Dots show location of consumers who do not buy the product from any of the stores. Black crosses show location of customers with $p_{jt} \leq 0.2$, dark gray crosses customers with $0.2 < p_{jt} \leq 0.3$, and light gray crosses customers with $p_{jt} > 0.3$.

The pattern of prices reduces the total quantity of product bought Q (see Table 1). This effect is due to the fact that stores in less densely populated areas prefer charging higher prices even at the cost of losing some of the potential customers (see location of the dotted consumers in Figure 1). The effect is negative and significant for all partitions of the data. The pattern of prices also affects the number and location of stores in the market. Competition reduces prices and profit margins of stores in larger and more densely populated settlements. Therefore, each individual store needs to serve larger market in order to cover its quasi-fixed cost. On the other hand, the number of stores in less densely populated areas might increase because the prices charged here are higher than under uniform pricing. However, markets in smaller settlements often cannot support more stores. This seems to be the reason why local pricing reduces the number of stores M for all the local pricing strategies for the entire dataset and for $p_R = 0.5$ and $p_R = 1$ (see Table 1). Furthermore, the effect of local pricing on the number of stores is negative and highly significant for all the remaining partitions of the data. Lower number of stores implies higher distance D, which increases the total shopping cost cD^2 (see Table 1). Again, this effect of local pricing on shopping cost is positive and highly significant for all the remaining partitions of the data.

The effect of local pricing on welfare W is negative and highly significant for the entire dataset as well as for all the partitions of the data (see Table 1). The change in welfare is $\Delta W = p_R \Delta Q - c\Delta D^2 - \Delta MF$. Welfare decreases because the negative effect of lower welfare due to lower quantity traded ($p_R \Delta Q < 0$) and higher distance ($-c\Delta D^2 < 0$) outweighs the positive effect of lower number of stores and therefore lower total quasi-fixed cost paid by all the stores ($-\Delta MF > 0$). Table 2 shows the contributions of the individual effects to the change in welfare for $p_R = 0.5$ and $p_R = 1$. Interestingly, the absolute value of ΔW is lower for low reservation price. It is because the consumer surplus from buying in supermarkets is low to start with, so the increase of their distance d^*_{jt} due to local pricing forces some of the consumers out of the market (see Figure 1). Which means that they do not waste as much welfare on shopping cost as consumers with high reservation prices.

reservation price	pricing strategy	ΔW	$\Delta Q p_R$	$-c\Delta D^2$	$-\Delta MF$
$p_R = 0.5$	$(\underline{L}, \underline{L})$	−17.8	−22.8	−17.8	22.8
	$(\hat{L}, \hat{L})$	−9.3	−7.7	−12.2	10.6
	(L_L, L_L)	−5.6	−9.3	−7.6	11.3
$p_R = 1$	$(\underline{L}, \underline{L})$	−40.4	−10	−52.9	22.5
	$(\hat{L}, \hat{L})$	−26.0	−3.0	−28.3	5.3
	(L_L, L_L)	−20.8	−2.9	−20.5	2.6

Table 2 The welfare effects of local pricing

And finally, Table 1 shows that the effect of local pricing on profits Π and consumers' surplus CS is ambiguous. It is due to the fact, that local pricing may lead to lower price P and revenue R, typically if the reservation price is low, or to higher price P and higher revenue R, if the reservation price is high. The intuition behind this result is as straightforward. If the reservation price is high relative to the equilibrium price in the market with uniform prices (e.g. $p_R = 1$ and $P = 0.307$), stores in areas with less competition are able to increase their prices substantially, which may increase also the average price P and revenue R. On the other hand, stores with relatively low reservation prices ($p_R = 0.5$) have less market power, therefore the average price P and the revenue R in the market are more likely to decline.

4 Conclusion

The aim of this paper was to investigate the effect of local pricing on market outcomes in the supermarket sector. For this purpose, the paper introduces an agent-based model with endogenous entry and location of stores. It finds that local pricing reduces the number of stores and the number of products sold in the market and increases the total shopping cost incurred by consumers. And since the welfare loss due to lower quantity sold and higher shopping cost outweighs the gain due to lower aggregate quasi-fixed costs, local pricing reduces total welfare. Similarly to [3], the effect on profit and consumers' surplus is ambiguous, but it is more likely that local pricing reduces the consumers' surplus. But differently to their model, the main factor that influences the direction of this effect in this model is the size of reservation price relative to the equilibrium price under uniform pricing.

Acknowledgements

This paper was created as a part of the specific research project no. MUNI/A/0796/2011 at Masaryk University.

References

[1] Borenstein, S.: Price discrimination in free-entry markets. *RAND Journal of Economics*, **16** (1985), 380–397.

[2] Dobson, P. W., and Waterson, M.: Chain-Store Pricing across Local Markets. *Journal of Economics & Management Strategy*, **14** (2005), 93–119.

[3] Dobson, P. W., and Waterson, M.: Price Flexing and Chain-Store Competition. *Proceedings of the 32nd EARIE Annual Conference*, (2005) available at http://www.fep.up.pt/conferences/earie2005/cd_rom/Session%20VI/VI.J/Dobson.pdf

[4] Dobson, P. W.: Relationship between Buyer and Seller Power in Retailing: UK Supermarkets, In: *Cases in European Competition Policy* (B. Lyons, ed.). Cambridge University Press, Cambridge, (2009), 100–128.

[5] Holmes, T. J.: The effects of third degree price discrimination in oligopoly. *American Economic Review*, **79** (1989), 244–250.

[6] Krčál, O.: Agent-Based Model of an Urban Retail Market. In *Proceedings of the 29th International Conference Mathematical Methods in Economics 2011*, (M. Dlouhý, and V. Skočdopolová, eds.), University of Economics, Faculty of Informatics and Statistics, Prague, 2011, 369–374.

Application of aging chain model on demographical data of the Czech Republic

Igor Krejčí[1], Roman Kvasnička[2]

Abstract. Paper deals with the design of dynamic aging chain model of Czech Republic. Population aging chains are necessary for different kinds of system dynamics models that focus on socioeconomic problems. Aging chain structure commonly consists of stock variables for age cohorts and flow variables for aging, births, deaths and migration. The basis of such computer simulation model that is based on system dynamics methodology is the system of nonlinear first-order differential equation.
In the first part, we will present three different aging chains quantified on the basis of current demographical data. Each chain is divided into two separate chains, one for each gender. Chains differ in the level of age groups aggregation and in the type of delay that is used to model the aging between age cohorts. Also two different approaches to represent changes in childbearing are explained. In second part, each simulation run is compared with official demographical data of the Czech Statistical Office and thus the best approach is identified on the basis of the accuracy of simulation.

Keywords: system dynamics, aging chain, population, fertility rate, delay.

JEL Classification: C63, C44, J12
AMS Classification: 91D20, 93C15

1 Introduction

Paper deals with the design of dynamic aging chain model of Czech Republic. Population aging chains are necessary for different kinds of system dynamics models that focus on socioeconomic problems. Aging chain structure commonly consists of stock variables for age cohorts and flow variables for aging, births, deaths and migration [1]. The core of such computer simulation model that is based on system dynamics methodology is the system of nonlinear first-order differential equation.

The aging chain structure is useful for non-human population simulation e.g. for housing [8] or capital in general [16]. Moreover similar structure to aging chain is possible use for model of persons' fluctuation between cohorts that are distinguished on basis of some more qualitative indicator e.g. work position promotion [12]. Nevertheless, we focused on human aging chain. Such chains are commonly used for macro models of whole world [13, 14] or country [17]. At the beginning, we set the goal to design versatile aging chain of Czech Republic population, which can serve as a molecule for more complex models [11]. The versatility should be ensured by disaggregation of age cohorts.

In the first part, we will present different aging chains quantified on the basis of current demographical data. Each chain is divided into two separate chains, one for each gender. Chains differ in the level of age groups aggregation and in the type of delay that is used to model the aging between age cohorts. Also two different approaches to represent changes in childbearing are explained. In second part, each simulation run is compared with official demographical data of the Czech Statistical Office and thus the best approach is identified on the basis of the accuracy of simulation.

2 Materials and methods

Aging chain is designed in Vensim DSS 5.10e. For parameters quantification the official demographical data from Czech Statistical Office were used. For fertility and initial stock values we used Czech demographic hand book [2], Demographic yearbooks of the Czech Republic [3, 4], for mortality parameters the life tables were necessary [5, 7].

[1] Czech University of Life Sciences Prague, Department of Systems of Engineering, Kamýcká 126, Prague 6, krejcii@pef.czu.cz.
[2] Czech University of Life Sciences Prague, Department of Systems of Engineering, Kamýcká 126, Prague 6, kvasnicka@pef.czu.cz.

Population model is divided according to gender, thus model contains two aging chains. Figure 1 present general stock and flow of aging chain, index *i* is age, index *s* stands for sex identification (*M* or *F*), box variables represent stock variables, pipes with arrow and faucet represent flow variables and simple arrows are causal links with polarity.

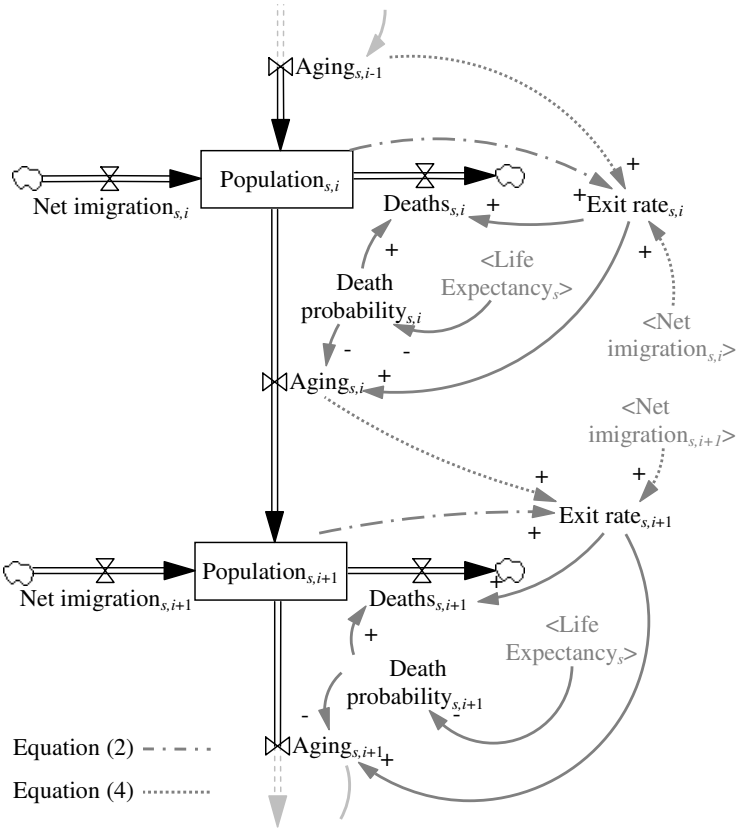


Figure 1 Simple stock and flow diagram of aging chain

In equation (1) stock variable population $P_{s,i}$ has two inflows and two outflows. $A_{s,i}$ is aging (maturation) from age cohort i to i+1 (thus inflow from previous cohort and outflow to following one), $I_{s,i}$ is net immigration and $D_{s,i}$, is outflow of deaths from population $P_{s,i}$. T is current time, initial time is T_0 and t is any time between T and T_0.

$$P_{s,i}(T) = \int_{T_0}^{T} \left(A_{s,i-1}(t) - A_{s,i}(t) + I_{s,i}(t) - D_{s,i}(t)\right)dt + P_{s,i}(T_0) \tag{1}$$

First aging chain in [11] used disaggregation into twelve age groups (i=0 was for age between 0–14 years, i=1, …, 10 is for of five-year groups, the last group i=11 is the population in the age of 65 and older). Next models used 72 age cohorts, one year cohorts until age of 70, then i=71 is for age 71 and more [10].

There are two possible ways to express exit rate $E_{s,i}$ (total outflow from population $P_{s,i}$). That is the reason why the causal link to exit rate missing in figure 2. Equation (2) is first order delay where $a_{s,i}$ is average time in cohort, thus in case with one year cohorts the equation (3) was used. First order delay generates exponential decay and assumes perfect mixing entities in stock [15].

$$E_{s,i} = \frac{P_{s,i}}{a_{s,i}} \tag{2}$$

$$E_{s,i} = P_{s,i} \tag{3}$$

Second possible approach is fixed delay. In this case the equation (4) express exit rate from $P_{i,s}$. Total output from population is total input to same population delayed by fixed time.

$$E_{s,i}(T) = A_{s,i-1}(T - a_{s,i}) + I_{s,i}(T - a_{s,i}) \tag{4}$$

The first approach in equation (3) is suitable in "situations where cohorts are defined not by age but by membership in category" [15] but in referred cases [13, 14, 17] it is used instead of equation (4) despite the cohorts are defined by age. Both types of delays were tested for exit rate in our models and appropriate causal links are shown in figure 1.

Anyhow the exit rate is defined; outflows of deaths and aging are obtained from equations (5) and (6). Death probability $DP_{s,i}$ is function of exogenous life expectancy, which is specified for both genders separately. We applied ordinary least square method on data from life tables [5, 7], the lowest value of determination coefficient R^2 is obtained for cohorts i=0, …, 15, it is always more than 0.75; for i=30, …, 71 R^2 is more than 0.9. In case of 12 age cohorts the R^2 is more than 0.9 for all cohorts.

$$D_{s,i} = E_{s,i} \cdot DP_{s,i} \tag{5}$$

$$A_{s,i} = E_{s,i} \cdot \left(1 - DP_{s,i}\right) \tag{6}$$

Official data on fertility focus only on females of age between 15 and 49 years (marginal cohorts are assumed to be 15 or less and 49 and more). The average age of mothers in the Czech Republic grew by more than two years in the first decade of 21st century. Thus, it was necessary to dynamise cohorts' fertility. We tested two approaches. In [10] we used equation (7) to calculate births B, FI_i is cohort fertility index and TF is total fertility rate. Again, the ordinary least square method was used; R^2 does not drop below 0.9. We used 20-year average ratio of male and female births, which means 51.4% of births are the inflow to male aging chain.

$$B = \sum_{i=15}^{49} P_{F,i} \cdot TF \cdot FI_i \tag{7}$$

Second approach uses density function of normal distribution. We defined standard deviation as function of average mothers' age with R^2 higher than 0.97. For extrapolation of exogenous variables we used estimates from [6]. In case of net immigration we used average age structure from years 2003-2010. Starting point of simulation is T_0 is 2003.5, i.e. 1st July 2003, fourth order Runge-Kutta integration with time step 0.03125 was used.

3 Results and discussion

The smallest model has more than 190 variables, the biggest more than eight hundred (144 stock variables, 432 flow variables of deaths, aging, migration and births, other variables are parameters – fertility indices, death probabilities, life expectancy etc.). We tested all mentioned combinations: five x one year cohorts, first order x fixed delay, fertility indices x normal distribution approximation.

Mean Absolute Percent Error was calculated for years 2004-2010, always for first of July. The worst results were obtained from aging chain with five year cohorts, first order delay and fertility calculated by (7). MAPE was 0.24 for whole population but it grows to more than 8.2 for female cohort of 30-34, which is very important for correct births estimation. Models with one year age cohorts produced better result. At first we tested aging chains that used (7). For both types of exit rate the worst results were for youngest age cohorts and showed possible improvement in estimation of fertility. The best results were obtained from aging chain that used fixed delay for exit rate (4) and normal distribution for fertility. Just 5 age cohorts has worse MAPE than the second best model with exit rate from (2) and also normal distribution approximation for fertility. Table 1 shows values of the mean absolute percent error MAPE for whole population, each gender and age cohorts.

System of demographic structure has high inertia, thus it is not surprising the MAPE for total population reaches always very good results for such short period. But the indicator doesn't fall if we use shorter period. More important is MAPE for specific age cohorts and, in this case, especially first six that are result of fertility modeling.

Age	Total	Male	Female	Age	Total	Male	Female	Age	Total	Male	Female
Total	0.164	0.132	0.194	**24**	0.340	0.515	0.157	**48**	0.475	0.597	0.350
0	0.979	1.233	0.714	**25**	0.515	0.573	0.452	**49**	0.358	0.479	0.234
1	0.500	0.376	0.658	**26**	0.613	0.723	0.495	**50**	0.354	0.433	0.273
2	0.636	0.533	0.744	**27**	0.431	0.468	0.393	**51**	0.290	0.314	0.297
3	0.519	0.483	0.557	**28**	0.462	0.394	0.534	**52**	0.273	0.321	0.225
4	0.488	0.530	0.443	**29**	0.469	0.447	0.494	**53**	0.161	0.184	0.162
5	0.330	0.408	0.270	**30**	0.589	0.646	0.529	**54**	0.144	0.164	0.164
6	0.220	0.219	0.221	**31**	0.507	0.537	0.474	**55**	0.128	0.149	0.108
7	0.175	0.235	0.152	**32**	0.490	0.536	0.442	**56**	0.110	0.152	0.085

8	0.181	0.168	0.195	**33**	0.338	0.365	0.310	**57**	0.101	0.131	0.073
9	0.091	0.108	0.107	**34**	0.447	0.543	0.346	**58**	0.058	0.075	0.061
10	0.106	0.097	0.135	**35**	0.365	0.387	0.341	**59**	0.107	0.164	0.055
11	0.067	0.065	0.085	**36**	0.404	0.490	0.314	**60**	0.088	0.163	0.05
12	0.141	0.143	0.157	**37**	0.455	0.571	0.332	**61**	0.169	0.252	0.094
13	0.078	0.103	0.101	**38**	0.522	0.579	0.463	**62**	0.139	0.235	0.062
14	0.086	0.077	0.112	**39**	0.379	0.442	0.313	**63**	0.258	0.403	0.132
15	0.158	0.134	0.183	**40**	0.410	0.528	0.286	**64**	0.217	0.354	0.105
16	0.145	0.135	0.154	**41**	0.185	0.308	0.079	**65**	0.346	0.453	0.256
17	0.468	0.371	0.581	**42**	0.304	0.410	0.192	**66**	0.217	0.303	0.184
18	1.239	1.404	1.067	**43**	0.244	0.329	0.155	**67**	0.302	0.378	0.241
19	1.900	2.371	1.408	**44**	0.357	0.397	0.316	**68**	0.226	0.253	0.238
20	1.911	2.380	1.459	**45**	0.279	0.436	0.125	**69**	0.377	0.460	0.314
21	1.098	1.558	0.648	**46**	0.352	0.439	0.262	**70**	0.359	0.421	0.312
22	0.615	0.834	0.397	**47**	0.312	0.474	0.148	**71+**	0.695	0.620	0.739
23	0.403	0.525	0.273								

Table 1 Mean Absolute Percent Error for best aging chain (%)

Figure 2 compares extrapolations from Eurostat [9], three extrapolations from the Czech statistical office [6] and results from most accurate aging chain model with fixed delay for aging and normal distribution for fertility. The presented result from modeled aging chain used low extrapolation of CZSO for exogenous variables, (i.e. total net immigration is 15,000 persons, LEF = 86.4 LEM = 81.6 in 2050) and middle extrapolation for fertility parameters (average mothers' age grows to 31 years in 2050 and total fertility grows to 1.72 in same year).

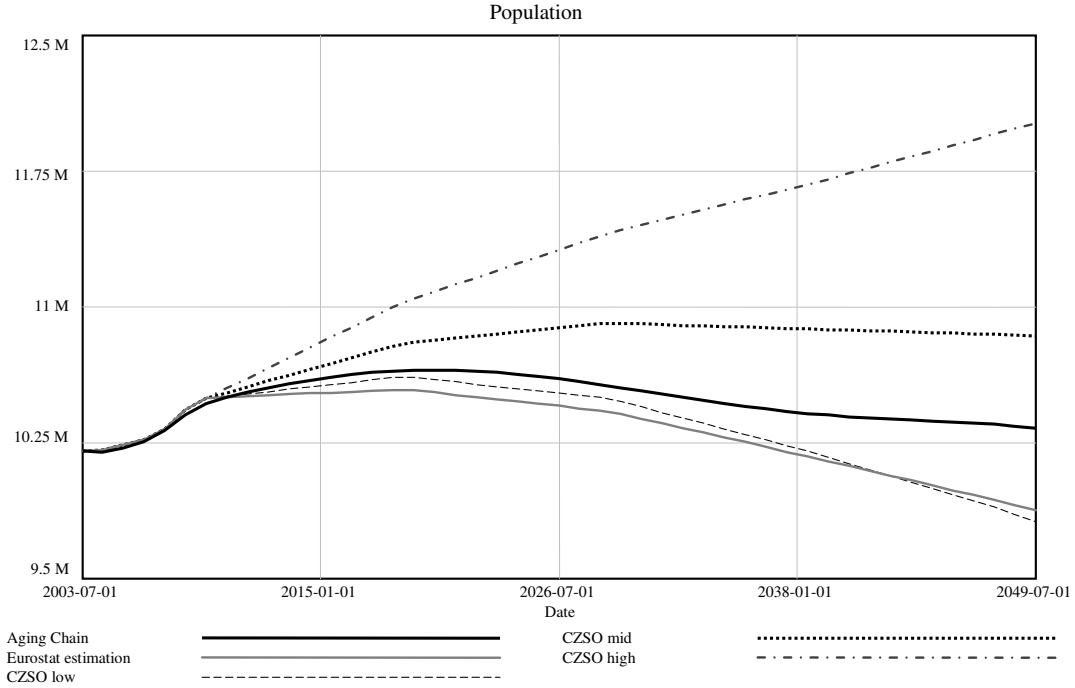


Figure 2 Simple stock and flow diagram of aging chain

The results shown that one-year cohorts are always better than five-year. Also the normal distribution approximation for fertility led to better results. Furthermore, the approximation simplifies the model construction (fertility indices require regression for every appropriate age cohort). Better results from fixed delay exit rate probably rise from highly non-uniform Czech population age structure, see figure 3 for comparison (cohort 71+ was excluded from the picture). On the left the structure is from second best model with first order delay like in [14] or [17], which smoothes the structure, on the right is the result from best model with fixed delay. The structure at 2003 is surveyed [2] and is same for both aging chains.

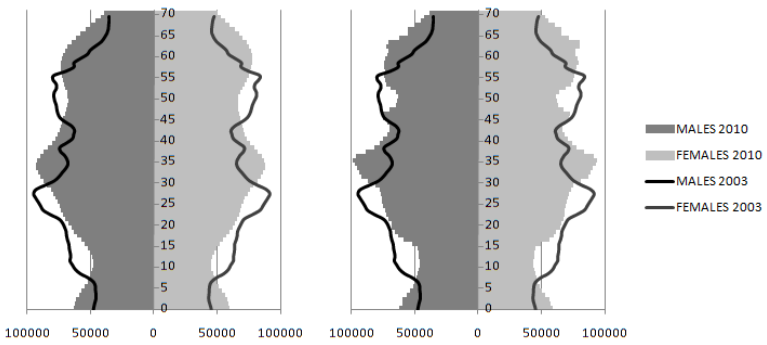


Figure 3 Comparison of age structure from different aging chains

4 Conclusions

We compared eight variants of aging chain with different age cohorts, different form of outflow from these cohorts and different approximation of fertility distribution. The best model used one year age cohorts, normal distribution approximation for fertility and fixed delay for outflow from age cohorts. The last is different from aging chains used in more complex models [14, 17], which rises from specific age structure of Czech population.

The presented aging chain is prepared for implementation into more complex system dynamics models. And we can provide it to any author for testing or implementation into their own model. Although it isn't constructed for purely predictive purposes it is also possible to use it this way.

Acknowledgements

The paper is supported by the grant project IGA FEM CULS 20121024 "Molecules of dynamic model of Czech republic".

References

[1] Bossel, H.: *System ZOO 1 Simulation models – Elementary systems, Physics, Engineering*. Books on Demand, Norderstedt, 2007.
[2] Czech Statistical Office: *Czech demographic handbook 2009*, [online] Czech Statistical Office, Prague, 2010, Available: http://www.czso.cz/csu/2010edicniplan.nsf/engp/4032-10 [20 Feb 2012]
[3] Czech Statistical Office: *Demographic Yearbook of the Czech Republic 2009*, [online] Czech Statistical Office, Prague, 2010, Available: http://www.czso.cz/csu/2010edicniplan.nsf/engp/4019-10 [15 Mar 2012]
[4] Czech Statistical Office: *Demographic Yearbook of the Czech Republic 2010*, [online] Czech Statistical Office, Prague, 2011, Available: http://www.czso.cz/csu/2011edicniplan.nsf/engp/4019-11 [15 Mar 2012]
[5] Czech Statistical Office: *Life Tables for the Czech Republic, Areas and Regions*, [online] Czech Statistical Office, Prague, 2010, Available: http://www.czso.cz/csu/2010edicniplan.nsf/engp/4002-10 [20 Mar 2012]
[6] Czech Statistical Office: *Projekce obyvatelstva české republiky do roku 2065* [Czech population projection up to 2065], [online] Czech Statistical Office, Prague, 2010, Available: http://www.czso.cz/csu/2009edicniplan.nsf/p/4020-09 [23 Mar 2012]
[7] Czech Statistical Office: *Úmrtnostní tabulky* [Life Tables], [online] Czech Statistical Office, Prague, 2010, Available: http://www.czso.cz/csu/redakce.nsf/i/umrtnostni_tabulky [23 Apr 2011]
[8] Eskinasi, M., Rouwette, E., Vennix, J.: Simulating urban transformation in Haaglanden, the Netherlands. *System Dynamics Review* **25** (2009), 182 – 206.
[9] Giannakouris, K.: *Ageing characterises the demographic perspectives of the European societies*, [online] Eurostat, Luxembourg; 2008, Available: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-SF-08-072/EN/KS-SF-08-072-EN.PDF [cit 10. May 2011].
[10] Krejčí, I., Kvasnička, R.: Modelling disaggregated aging chain of Czech population. In: *Proceedings of the 29th International Conference on Mathematical Methods in Economics 2011* (Dlouhý, M., Skočdopolová, V. eds.), 2011, 375-379.
[11] Krejčí, I, Kvasnička, R. Švasta, J.: Dynamic aging chain of the Czech Republic population. *Acta universitatis et Silviculturae Mendeleianae Brunensis* **59** (2011), 209-216.
[12] Meadows, D. et al.: *The Limits to Growth*, Universe Books, New York. 1972.
[13] Meadows, D. H., Randers, J., Meadows, D.: *The Limits to Growth: The 30-year update*. Chelsea Green Publishing Company, Vermont, 2004.

[14] Sterman, J. D.: *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Irwin/McGraw-Hill, Boston, 2000.
[15] Kunc, M.: Achieving a balanced organizational structure in professional services firms: some lessons from a modelling project. *System Dynamics Review* **24** (2008), 119 – 143.
[16] Wang, Q., Ma, S.: Some of the limits to growth of big city in China. In: *Proceedings of The 5th International Conference of the System Dynamics Society* (editoři eds.). System Dynamics Society, Shanghai, 1987, 615–628.
[17] Wang, Q., Sterman, J.: A disaggregate population model of China. *Simulation* **45** (1985), 7 – 14.

The impact of alternative approaches to the measurement of fixed capital

Igor Krejčí[1], Jaroslav Švasta[2]

Abstract. The most recommended method used for the measuring of fixed capital is PIM (Perpetual Inventory Method). In this method, the processes of retirement, decay and obsolescence of fixed assets in national economy are modelled on the basis of average service lives, and the chosen depreciation profile.
The paper deals with the comparison of two most common depreciation profiles. The straight line depreciation profile is used by official national accounts statistics. Having no data on second hand market or rental prices, the geometric profile is derived from average service lives. As an alternative, the geometric depreciation profile is applied on the data on most important types of assets for non-market producers sector. As a consequence, the impact on gross domestic product of such change is evaluated.
Also the gradual change of service lives was tested on assets with long service lives. Subtle impact of this process on balances of fixed assets and GDP was demonstrated.

Keywords: fixed capital, national accounts, depreciation profile.

JEL Classification: C44, E22
AMS Classification: 91B25

1 Introduction

It is nearly impossible to make the stock list and its evaluation in purchaser prices of all fixed assets in the economy. But the value of fixed capital is highly demanded indicator. Fixed capital in economy is represented by a few flows and stocks. Whereas stock variables (gross and net) are estimated to some point in time (end of year) flows refer to period of time (year).

The most recommended method [6, 7] used for the computation of stocks and the consumption of fixed capital is PIM (Perpetual Inventory Method). The core of the method lies in accumulation of flows to express the stock variables. While flow of investments is possible to survey, the values of depreciation and retirement have to be estimated. The processes of the retirement, decay and obsolescence of fixed assets in national economy are modelled on the basis of average service lives, and the chosen depreciation profile [11].

To estimate the process of consumption of fixed capital (i.e. decline of assets value during some period of time "… as a result of physical deterioration, normal obsolescence or normal accidental damage" [7]) it is necessary to determine the depreciation profile. The most used are straight line and geometric profiles.

Both profiles are criticised for some reasons. Straight line depreciation profile often doesn't correspond to real economy situations; it is too slow, especially for periods of dynamic structural shifts [1]. On the other hand, geometric depreciation is often criticised for too fast depreciation and geometrically depreciated assets never reach zero value [5, 8]. Moreover, empirical evidence for geometric depreciation based on second hand market prices is hardly representative [8], such criticism can be focused also on rental prices approach (see [8] and [9] for more about empirical evidence of geometric depreciation profile). Thus, the depreciation profile is frequently chosen on the grounds of practicality or expert's judgement.

Moreover average service lives change in time and should be revised periodically [12]. Some statistical offices also assume gradual change (mostly decrease) of service lives of long-life assets [11].

The paper deals with the comparison of these two most common depreciation profiles and with the implementation of gradual change of service lives. Czech official balances of fixed assets are based on straight line depreciation profile [2]. Alternative approaches are applied on official data of Czech statistical office (CZSO) and the impact on gross domestic product of such changes is evaluated. Geometric depreciation profile

[1] Czech University of Life Sciences Prague, Department of Systems of Engineering, Kamýcká 126, Prague 6, krejcii@pef.czu.cz.
[2] Czech University of Life Sciences Prague, Department of Systems of Engineering, Kamýcká 126, Prague 6, svasta@pef.czu.cz.

is applied twice for average service life as constant number (i.e. simultaneous exit) and for lognormal retirement pattern that is used at CZSO [2, 13].

2 Material and Methods

Perpetual Inventory Method is based on annual accumulation of flows (gross fixed capital formation, consumption of fixed capital, holding gain/loss and other changes). Consumption of fixed capital is modelled decrease of value of old investments. Thus, investment (gross fixed capital formation) is input flow in current year and also the cause of output flow of consumption of fixed capital for certain period of following years.

Under commonly used straight line depreciation profile, the value of assets decrease by constant amount each year. Thus value loss of assets follows equation (1) where T is average service life, n is age of asset, p_0 is initial value of investment and p_n is actual value of the assets after n years [11]:

$$\frac{p_n}{p_0} = 1 - \frac{n}{T},\ n = 0,1,...,T \tag{1}$$

Value loss of asset under geometric depreciation profile is expressed by (2).

$$\frac{p_n}{p_0} = (1-\delta)^n,\ n = 0,1,... \tag{2}$$

Depreciation rate δ should be based on empirical data [11] but in case of no econometric estimates on δ, it is possible to convert assets service lives into depreciation rates. Equation (3) dependence of δ on T [11]. Under assumption of long-run equilibrium for straight line and geometric depreciation (double declining balance) [5]:

- long-run constant investment in constant price,
- long-run value of the capital is equal for both depreciation profiles.

the declining-balance rate R could be chosen equal to 2 [5]. Empirical results differ from that value in both directions – some authors find R smaller than 2, some higher [11]. This assumption (R=2) was used for the results presented in this paper.

$$\delta = \frac{R}{T+1} \tag{3}$$

Figure 1 compares different depreciation profiles on example of 100 million CZK investment into fixed assets with average service life T=5 in year 0. Horizontal axis represents age; vertical is the value of assets. Figure 1 also express importance of retirement function. Simultaneous exit assumes all assets to be retired at age T. Recommended bell-shaped function [11] is represented by lognormal retirement function with standard deviation of service life s=2.9. Such retirement functions divide each vintage of assets into groups with different service lives.

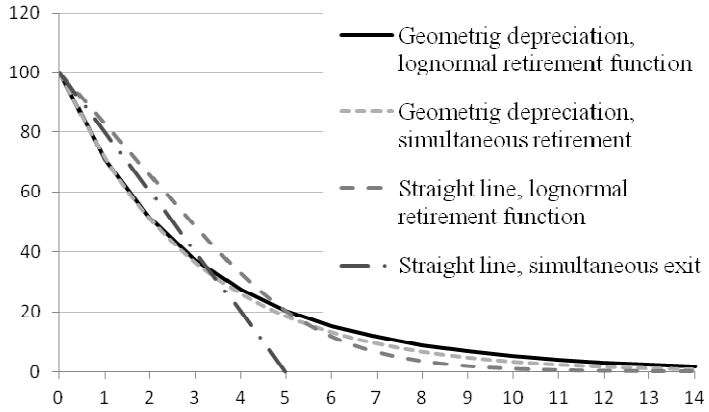


Figure 1 Value loss in time – example of different depreciation profiles, million CZK, prices of basic year

The impact is obvious especially for straight line depreciation. Value loss under straight line depreciation with lognormal retirement function is closer to geometric depreciation than to straight line with simultaneous

exit retirement pattern. Straight line depreciation profile and lognormal retirement function are typical for CZSO [2, 4], thus there shouldn't be expected such change as it was for official revision of national accounts, which part was change from simultaneous exit to lognormal retirement function for non-residential buildings and application of PIM on dwellings instead of price-quantity method [4].

Figure 1 also shows why it isn't clear in which direction will the consumption of fixed capital and value of net capital stock change. The result depends on historical time series of gross fixed capital formation. Faster depreciation of historical investments could lead to smaller consumption of fixed capital in current period.

Alternative approach was applied on data on government sector and institutional sector of non-profit organisations serving households. These are institutional sectors of non-market producers where the consumption of fixed capital is necessary for GDP estimation [7]. Geometric depreciation profile was applied on following types of assets:

- Non-residential buildings;
- Dwellings;
- Transport equipment;
- Other machinery and equipment.

These kinds of assets are often subject of discussion [5, 8] and in 2010 they represented more than 86.5% of value fixed assets in Czech economy [3]. Impact of 1% decrease of T was tested for non-residential buildings and dwellings; for these purposes standard deviation for lognormal retirement pattern was set in the middle of recommended interval $\langle T/4, T/2\rangle$ [11]. For practical purposes the maximal service life is set on 99.5 percentile of original T. Service lives started to fall in 2002 for dwellings and in 2003 for non-residential buildings. For these years the official PIM is parameterised.

Calculation respects the format of data input, thus it is classified into 120 industries (average service lives and price indices often differ for different industries) and the alternative model uses transformation to Markov chain based on [10]. Data for straight line depreciation are from CZSO [3]. For year 2009 it is semi-definitive version, for 2010 it is just preliminary version of data.

3 Results and discussion

Decreasing average service lives for chosen long-life assets caused very small changes. The value of net capital stock in current prices decreased just by 0.49‰ for non-residential buildings and by 0.57‰ for dwellings in 2010 (i.e. by 1,742 billion CZK). Consumption of fixed capital in current prices increased by 0.36% for non-residential buildings and by 0.31% for dwellings for the same year (i.e. by 281 million CZK).

Geometric depreciation profile caused changes in both directions. Figure 2 compares development of consumption of fixed capital in current prices based on different depreciation profiles. Until 2004 geometric depreciation provides higher estimation of consumption of fixed capital for non-market producers from 2005 the consumption is below official linear profile (data from [3]). Consumption of fixed capital continuously grows only for transport equipment.

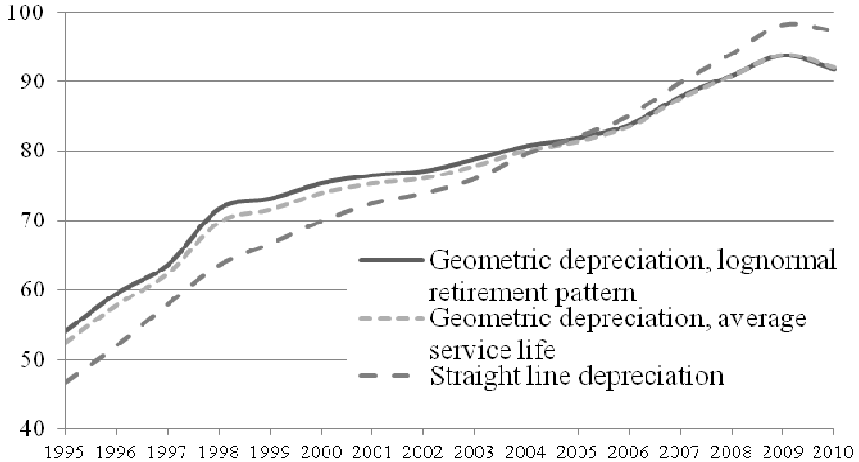


Figure 2 Non-market producers' consumption of chosen assets in billion CZK, current prices.

The biggest positive difference from official data is for 1995. It is caused mainly by non-residential buildings 86.6% of total difference for lognormal retirement function and 87.8% of the difference for simultaneous exit.

The decrease is caused mainly by decreasing consumption of dwellings and non-residential buildings. Consumption of dwellings in sector of non-market producers calculated on basis of geometric depreciation profile is under 70% of official statistics for simultaneous exit in 2010 and fewer than 60% for lognormal retirement pattern (i.e. drop by 2,414 and 2,944 million CZK in current prices).

In case of simultaneous exit, consumption of non-residential buildings dropped by 4.4% (3,118 million CZK) in 2010. For lognormal retirement function it dropped by 4.3% (3,058 million CZK). But these changes can't be generalised for whole economy. For whole economy the consumption of dwellings is higher than 92.2% for both retirement patterns in 2010. For non-residential buildings in Czech economy the consumption of fixed capital is even higher than official in 2010 (4.6% for simultaneous exit and 6.8% for lognormal retirement).

Assumption of simultaneous exit is typical for many researchers and statistical offices [5]. The difference of geometric consumption between simultaneous exit and lognormal retirement pattern is always under 3.2%. Although this result is hard to generalize, for practical purposes the computational complexity should be considered. For long-life assets (e.g. non-residential buildings) there could be more than 200 times more computations for each year of investment.

The consumption of fixed capital is necessary for calculation of production of non-market producers. Thus the estimation of their consumption influences GDP. Table 1 presents impact of different depreciation patterns on GDP of Czech republic. There is also the market production in government institutional sector. The impact on GDP presented in table 1 is caused only by consumption of fixed capital related to non-market production. In comparison, already mentioned change in consumption of fixed capital during revision of national accounts caused changes of GDP between 0.0% and 0.3% in 1995-2009 [4].

retirement pattern	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Δ lognormal	7 006	6 601	4 324	6 209	4 441	3 275	2 037	1 608	1 607	219	-809	-1 571	-1 992	-2 897	-3 855	-4 509
Δ simultaneous exit	5 425	4 964	3 097	4 364	2 908	1 922	877	524	513	-662	-1 508	-2 165	-2 583	-3 347	-4 155	-4 680
Δ lognormal (%)	0.46	0.37	0.23	0.30	0.21	0.14	0.08	0.06	0.06	0.01	-0.03	-0.05	-0.05	-0.08	-0.10	-0.12
Δ simultaneous exit (%)	0.35	0.28	0.16	0.21	0.14	0.08	0.04	0.02	0.02	-0.02	-0.05	-0.06	-0.07	-0.09	-0.11	-0.12

Table 1 The impact of geometric depreciation profile on GDP estimation, million CZK in current prices and relative change

4 Conclusion

The change of GDP under 0.5% isn't high, especially for actual period where it is 0.12% decrease. Moreover, results could be different for different R in equation (3) and also higher detail of input data (official PIM uses classification on institutional subsectors [2]) can slightly change the result. But concerning criticism of tested depreciation patterns (first is too slow, second is too fast), application of both depreciation profiles can be comprehended as interval estimation of consumption of fixed capital and other associated indicators.

Future work in this area will focus on the comparison for whole Czech economy. Concerning the change of net fixed capital stock that is -13.9% for tested sectors in 2010, the impact on factors' productivity (see e.g. [14]) will be also in the area of interest and detailed examination.

Acknowledgements

The paper is supported by the grant project IGA FEM CULS 20121024 "Molecules of dynamic model of Czech republic".

References

[1] Chirinko, R. S. Business Fixed Investment Spending: Modleing Strategies, Empircal Results, and Policy Implications. *Journal of Economic Literature* **31** (1993), 1875-1911.

[2] Czech Statistical Office: *Gross national income inventory*. [online] Czech Statistical Office, Prague, 2002 Available: http://apl.czso.cz/nufile/GNI_CZ_en.pdf [11 Mach 2012].

[3] Czech Statistical Office: *Annual national accounts*. [online] Czech Statistical Office, Prague, 2012 Available: http://apl.czso.cz/pll/rocenka/rocenka.indexnu_en [20 Mach 2012]

[4] Czech Statistical Office: *A commentary to results of the extraordinary revision of Annual national accounts of the czech republic 1995-2009*. [online] Czech Statistical Office, Prague, 2012 Available: http://apl.czso.cz/nufile/RevizeRNU2011en.pdf [25 Mach 2012]
[5] Diewert, W. E.: Issues in the Measurement of Capital Services, Depreciation, Asset Price Changes and Interest Rates. In: *Measuring Capital in the New Economy* (Corrado, C., Haltiwanger, J., Sichel, D. eds.). University of Chicago Press, Chicago, 2005, 479-542. ISBN 0-226-11612-3.
[6] Eurostat, United Nations, International Monetary Fund, OECD, World Bank: *System of National Accounts 1993*, New York. 1993.
[7] Eurostat, United Nations, International Monetary Fund, OECD, World Bank: *System of National Accounts 2008*, New York. 2009.
[8] Hulten, Ch., R., Wykoff, F., C.: Issues in the measurement of economic depreciation introductory remarks, *Economic Inquiry* **34** (1996), 10-23.
[9] Jorgenson, D. W.: Empirical Studies of Depreciation, *Economic Inquiry* **34** (1996), 24-42.
[10] Krejčí, I.: Computation of Stocks and Consumption of Fixed Capital with use of Markov Chain. In: *Proceedings of the 28th International Conference on Mathematical methods in Economics* (Houda, M., Friebelová, J. eds.). University of south Bohemia in České Budějovice, České Budějovice, 2010, 365–370.
[11] OECD: *Measuring Capital - OECD Manual 2009*: *Second edition*, OECD Publishing, Paris, 2009.
[12] Schmalwasser, O., Schidlowski, M.: Kapitalstockrechnung in Deutschland. *Wirtschaft und Statistik* **11** (2006) 1107–1123.
[13] Sixta, J.: Odhady spotřeby fixního kapitálu. *Statistika* **87** (2007), 156 -163.
[14] Sixta, J., Vltavská, K., Zbranek, J. Souhrnná produktivita faktorů založená na službách práce a kapitálu. *Politická ekonomie* **59** (2011), 599-617.

Allocation of trains to platforms optimization

Michal Krempl[1]

Abstract. An important part of a transport market is a railway transport. A train traffic diagram is a key element of an operational implementation of a railway traffic. A track occupancy plan is its annex, which specifies a schedule of trains staying on tracks by platforms in the particular railway station. A track occupancy plan is an important tool for the dispatcher helping him allocate trains to platforms. Unfortunately in the Czech and Slovak Republic planning train movements within the station is done by hand, using planner's experience and a set of rules determined by a railway company. This article deals with constructing a track occupancy plan in the railway station using modern computational methods for its automatic creation based on the entered input data (the daily timetable and the structural and operational constraints). The linear mathematical model is introduced, the meaningful optimization criterion is designed and the constraints are formulated. The problem of allocation of trains to platforms is formulated as a single criteria optimization problem. The objective function minimizes the total occupancy time of a railway station infrastructure.

Keywords: transport, optimization, train platforming, railway stations, linear mathematical modeling.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Railway stations are the basic elements of railway infrastructure. Stations are the most fundamental and complex element of infrastructure in terms of rail transport organization. Train traffic control is performed according to the principles of one executive dispatcher within railway station or within remote control track sections (including rail stations).

Dispatcher is a worker realizing operational management. This type of control is characterized by long, continuous and repetitive decisions depending on the current situation. Dispatcher has to decide on these basic processes:

1. the organization of train traffic:
   a) a decision on a path for train arrival, departure and transit or reject such a request, in other words: decision to allocate a track for incomig and leaving train, or reject this request,
   b) a decision to change the sequence of trains
   c) a decision to wait for connecting trains
2. the organization of shunting
   a) a decision on the status of shunting paths, or rejection of this request

All decisions are done in relation to time when the decision is made, and in relation to the current operating situation.

Now, we focus on decision-making process referred in point 1) and only for the case of passenger trains arriving and only for the train. Dispatcher determines "only" the rail (the platform) the train is about to arrive when deciding on the implementation of the specified process.

This process is critical for the railway transport organization. Decision-making process is very complicated because it is necessary to take into consideration several parameters, if we demand that the decision was good. These criteria can be divided for greater clarity in the following categories:

- traffic (possibility of infrastructure due to the planned train paths, the length of the platform due to the length of the train, train direction, changing the sequence of trains, time interval between train going in the same direction, the speed of trains going in in the same direction, train categories, the degree of influence of trains going in opposite direction, lock-out or failure of the infrastructure elements - rails, switches, signals, bridges, etc., current and future station tracks occupation),

[1] VŠB - Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, Ostrava-Poruba, 708 33, Czech Republic, michal.krempl.st@vsb.cz

- transportation (quality of platform equipment: lifts, information boards, height of the platform edges, wheelchair accessibility, lighting, subway access, protection from the weather, distance to the connecting trains, etc.).

In general, you can think of two basic types of decision support tools for the dispatcher's decisions on the train traffic organization:

- static (track occupancy plan), taking into account deterministic conditions, which are constant during the planning period of train traffic diagram (TTD). This not reflects the current operating situation - deviations from TTD.
- dynamic, reflecting deviations from TTD. Dynamic tools are not used in the Czech Republic now.

Track occupancy plan is usually compiled for stations with a wide range of rail passenger transport and with complex transport infrastructure which is characterized by high or low number of platforms (platform edges). The plan is usually compiled in one year period - for the duration of TTD. The plan is based on deterministic (non-random), conditions.

Currently the plan is compiled in the Czech Republic without the use of sophisticated computational methods (i.e. simulation, operations research, graph theory, etc.). The plan is done "manually" with the support of a computer (software) display of the continuous working and the final draft. The aim of the plan making process is to obtain the optimal variant, but its attainment is not guaranteed. Value of the final solution is determined by the quality (theoretical knowledge, practical experience and personal qualities) of employees compiling it. The plan done in several basic steps:

1. Preparation of input data from the timetable draft for a particular station, the station order and related regulations,
2. Making acceptable plan while trying to obtain the best possible solution. If you cannot find an acceptable solution, feedback to the timetable draft is realized (change the time position of trains, cancellation of train paths, etc.)
3. Increasing value by permissible sub-plan modifications,
4. Final check of the founded solution admissibility (theoretical verification of its feasibility) and the feedback to TTD.

What opportunities are opening up finding a better quality plan? First, it is appropriate to note, in what cases it can be very beneficial to find a better quality of plan and when the benefits will be probably minimal.

Good plan should have impact for improving the train traffic organization quality in the following areas:

1. improving railway transport quality in passenger‘s angle of view:
   a) reduce train change distances,
   b) reduce the probability of platform changes due to the schedule,
   c) arrival and departure of trains to and from more appropriate platform (with sufficient equipment).
2. improving railway transport quality in infrastructure manager‘s angle of view:
   a) shorten the average time when incoming and departing train occupies a path in railway station,
   b) uniform use of infrastructure elements (track, platforms), etc.
3. acceptation of carriers the requirements:
   a) allocate the train to required time position,
   b) allocate the train to the requested platform and specific track in the railway station.

2 Problem formulation

Now, let us concentrate on compilation of the optimal track occupancy plan in the railway station. Operations research methods - linear mathematical modeling are used to obtain the optimal one.

The following basic steps should be done while compiling the mathematical model:

1. Appropriate choice of optimization criterion (criteria).
2. Definition of decision to be made and their expression by decision variables, including the determination of range of variables (obligatory conditions).
3. Formulation of the objective function.
4. Formulation of constraints and their mathematical expression using linear relationships between variables,
5. Modification of the model to reduce computational demands on solving methods.

2.1 Choice of optimization criteria

Earlier, we outlined two basic perspectives on the organization of rail transport - the allocation of the station tracks for incoming train:

a) traffic (operational) view,

b) transportation (passenger view).

We can define the following optimization criteria and the direction of optimization accordingly to the above points of view:

a) to minimize the total occupancy time of each individual elements of infrastructure (tracks, switches).

The aim is to choose such a train path from the entrance light to the platform that the occupation period of rails, and nodal points will be the shortest. The purpose of this criterion is to build a robust plan that would minimize the security risk and also the degree of fault tolerance. The shorter is the track occupied (i.e. the fewer nodal point is in the train path), the lower is the possibility of collision with another train. Therefore the probability that the delay of one train causes a delay of other trains will be reduced. For the same reasons it is desirable to minimize the period when train stay by the platform and give a train some station track which is not too busy in order to maximize a free time after departure of a previous train. While using this criterion we can also consider using this mathematical model to calculate the capacity of the railway station - to determine the maximum number of trains that can station serves at a specified time.

b) to maximize passenger satisfaction, which includes two main sub-criteria:

i. minimize train change distances. This is the total distance that each passenger is forced to overcome when changing train or to exit station.
ii. maximize of train to platform assignment fitness in terms of platform equipment.

c) to maximize the satisfaction of the carrier (carriers)

i. minimize deviations from the train time position,
ii. minimize deviations from the desired track by platform at a railway station.

The optimization criterion we define as follows: the total occupancy time of each individual elements of rail infrastructure while riding the train. Optimization direction: minimization.

3 Definition of decision variables and their domains

Previously, we have been specified key decision-making on the allocation of the incoming train to station tracks including criteria that dispatcher usually takes into account. In practice, decision-making process usually takes place in the following steps:

1. Decision to accept the train to the railway station, (the choice of yes / no).
2. Decision on making the train path from the entrance light to the platform, (the choice of yes / no).
3. Selection of a suitable path for approaching train - train platform allocation, (choice of at least one option)
4. Making train path and the entrance lights to the position allowing the drive, (no choice, it was already done in step 3),
5. Deregistration to the front station and the abolition of a train path (choice of yes / no).

The purpose of performing steps 1-2 and 4-5 is to find acceptable solution in the allocation of the incoming train to station tracks. Step 3 is "optimizing", the dispatcher tries to get some solution by the available methods which is at least far from optimal. In the implementation of step 3, in terms of the chosen optimization criterion, the dispatcher is looking for a train path which is occupied the shortest time. This means that the sum each individual elements of infrastructure (tracks, switches) occupancy time is minimal. With some simplification one can say that we are looking for a path that will occupied by arriving train the shortest time, will be released quickly and ready for another train. When applying this approach to find the optimal plan cast track is expected to increase the number of trains that can be served in a given time period in the railway station. Primarily, the requirements of carriers and passengers are not taken into account. Although, as indicated in the "restrictive conditions", any limitation can be define in the allocation of train to platform, eg:

- T84 train must come to the platform P4 or P5,
- trains from the direction of D2 cannot come to the platforms P1, P2, P7, etc.

In guidance line with previous text, we define variable x_{ikt}. This is a bivalent variable, domain is set $\{0,1\}$. The variable is used to model the decision whether train i occupies infrastructure element k in time t. If the value of a variable x_{ikt} is 1, train i occupies infrastructure element k in time t. Otherwise x_{ikt} variable takes value 0. We define the sets and their elements:

- $i \in I$ ... element of the set of all trains that arrive at the railway station,
- $k \in K$... the element of the set of all elements of the infrastructure in the railway station
  - $n \in N$... element of the set of all platforms (or tracks at the platform edge) in the station,
  - $p \in P$... element of the set of tracks and switches in the train station,

the following applies: $K \cong N \cup P$; $N \cap P \cong \{\ \}$

- $t \in T$ ... element of the set of time.

3.1 Formulation of the objective function

The basic type of decision to be made has been defined in the previous section. Now, let's qualify that decision, If we have in mind the chosen optimization criterion, assign each item of infrastructure the time (time interval) during which it is occupied when passing train, and denote this value h_{inp}.

- h_{inp} ... time during which the train i incoming to the platform n occupies infrastructure element p.

Platform occupancy time is given by TTD, which is specified by length of stay of each train in the railway station. The period of tracks and switches occupancy is on the decision of the dispatcher. Length of this period depends on track allocation for the approaching train. This time is affected by these factors:

- speed of a train,
- direction of the train (from which the previous station to which following station train rides).

These two factors determine the routing of the train at the railway station:

- nodal points, which are occupied by a train while driving,
- tracks, which are occupied by a train while driving.

Objective function takes the form:

$$\min\ f(x) = \sum_i \sum_k \sum_t h_{inp} \cdot x_{ikt} \tag{1}$$

3.2 Formulation of the restrictive conditions

So far formulated the mathematical model ensures that the obtained solution is optimal. But there is no guarantee that this solution will be realizable in practice. For this reason, it is necessary to define requirements, we have to address in terms of its validity.

A. Each train will come to the railway station just once.

B. Each element of the infrastructure is occupied by at most than one train at one time.
   a. Each track at the platform is occupied by at most than one train at one time.
   b. Each track off the platform and switch is occupied by at most one train at one time.

C. All the relevant elements of infrastructure are occupied in terms of selected path for the train.

D. Each element of infrastructure for the selected path and the train is occupied for given period, which is:
   a. sum of direct and indirect occupation of track for the platforms and tracks at the platform edges. Time of direct occupation is period of train staying in the station referred to TTD,
   b. time referred to the technical possibilities of station safety devices.

Formulation and completion of the requirements A-D, guaranteed that the obtained solution will be valid and also realizable in practice.

Now let's formulate constraints using mathematical notation by declared variables.

$$\sum_{k \in N} \sum_t x_{ikt} = S_i \quad for\ \forall i \in I \tag{2}$$

Wording conditions (2) is as follows: For each train i is assigned the platform n just to stay in the station for the time S_i.

Condition (2) guarantees the fulfillment of conditions A, D.a. Conditions (2) ensures that every train i arrive to the railway station just once and occupy the platform n just to stay on the track for the specified time. It cannot therefore be the case that the train i is assigned to two different platforms. It is precisely defined period of occupancy of one of the elements of infrastructure - platforms. The exact definition of the occupancy time of other infrastructure elements (switches and tracks off the platforms) will be carried out under condition (4).

$$g_{in} \cdot x_{\text{int}} \le \sum_{k | h_{inp} > 0} x_{ikt} \cdot h_{ink} \quad for\ \forall i \in I, n \in N, \forall t \in T \tag{3}$$

- g_{in} ... the sum of occupancy time of each individual element of infrastructure which are occupied when the train i is assigned to platform n
- x_{int} ... bivalent variable modeling the decision whether the train i occupies the platform n at time t

- x_{ikt} ... bivalent variable modeling decision if the train i occupies the infrastructure element k at time t
- h_{ink} ... time during which the train i going to the platform n occupies the element of the infrastructure k

Wording conditions (3) is as follows: If the train i come to the platform n at time t, during the time period between making and abolition of the train path are occupied tracks and switches off the platform corresponding to the selected path for the train i to the platform n. Condition (3) guarantees the fulfillment of condition C. Condition (3) also checks whether the route for the train i was built on time.

Condition (3) does not guarantee that each of the infrastructure elements will be busy for a corresponding period. It is therefore necessary to define a new condition that will ensure compliance of this requirement.

$$\sum_{f_k}^{f_k+h_{inp}-1} x_{ipt} \geq h_{inp} \cdot x_{\text{int}} \quad for\ \forall i \in I, n \in N, p \in P \tag{4}$$

Wording conditions (4) is as follows: If train i is assigned to platform n with the beginning occupancy time f_i, then from time f_k to time $f_k+h_{inp}-1$ are occupied switches and tracks off the platforms corresponding to the selected path for the train i going to platform n. There are all variants tested in combination train - platform – switches - track off the platform. If $h_{inp} = 0$ (path via given element p cannot be build), then the condition is always satisfied and minimization objective function causes that no element of infrastructure is busy.

Condition (4) ensures the condition D.b is fulfilled. Condition (4) checks for each element of infrastructure, whether it is occupied since making the train path for the right time.

Compliance with the conditions B.a., B.b. is given by the variable domain x_{ikt}. X_{ikt} variable can take a maximum value of 1, occupation of infrastructure elements by more than one train is rejected in any case.

Final form of the mathematical model follows:

$$\min\ f(x) = \sum_i \sum_k \sum_t h_{inp} \cdot x_{ikt} \tag{1}$$

$$\sum_{k \in N} \sum_t x_{ikt} = S_i \quad for\ \forall i \in I \tag{2}$$

$$g_{in} \cdot x_{\text{int}} \leq \sum_{k | h_{inp} > 0} x_{ikt} \cdot h_{ink} \quad for\ \forall i \in I, n \in N, \forall t \in T \tag{3}$$

$$\sum_{f_k}^{f_k+h_{inp}-1} x_{ipt} \geq h_{inp} \cdot x_{\text{int}} \quad for\ \forall i \in I, n \in N, p \in P \tag{4}$$

$$x_{ikt} \in \{0,1\} \tag{5}$$

4 Input data

Dates about the movement of trains are important when compiling robust plans, particularly deviations from the timetable.

Following measurement were done to find delays at railway stations in order to analyze the input data for the mathematical model:

- train arrival to Prague main station in the period from April 10, 2011 19:00 to April 11, 2011 6:00. Statistical set contains 97 observations. All measured data $\epsilon \langle -9,20 \rangle$,
- train arrival to Prague main station in the period from April 15, 2011 to April 16, 2011 19:00 6:00. Statistical set contains 83 observations. All measured data $\epsilon \langle -21,34 \rangle$,
- train arrival to Brno main station in the period from May 06, 2011 22:00 to May 07, 2011 6:00. Statistical population represents 33 observations. All measured data $\epsilon \langle 0,25 \rangle$,
- train leaving from Brno main station in the period from May 06, 2011 22:00 to May 07, 2011 6:00. Statistical set constraints 35 observations. All measured data $\epsilon \langle 0,25 \rangle$.

For all the statistical observation was made χ^2 test with the following results:

- In the case of Prague Main Station we do not reject the hypothesis that the measured data come from a normal probability distribution.
- In the case of station Brno main station we do not reject the hypothesis that the measured data come from an exponential probability distribution.

We also carried out further statistical tests. Results of these tests have not challenged the conclusions resulting from the χ^2 test.

We say that the train runs deviations from the timetable with the support of statistical tests carried out – specifically the arrival to the railway station Prague m.s. is governed by the normal probability distribution with parameter: $EX = \mu = 0{,}42$. The dispersion in the data sample was estimated as follows: $DX_1 = 22{,}75$; $DX_2 = 41{,}91$.

5 Conclusions

In the paper an integer programming model for routing and scheduling trains at a passenger railway station is described. The model gives an optimal solution with regard to chosen criteria, using a general MP solver, the solution for a large station can be found in a reasonable time of a few minutes.

Now I focus on meeting real input data from a particular railway station. Another goal is to find an optimal track occupancy plan in term of selected criterion.

References

[1] Bažant, Michael: Řešení vybraných provozních problému osobní železniční stanice v rámci simulačního modelu. *Pardubice, 2009. 95 p. Univerzita Pardubice.*

[2] Bažant, Michael, Kavička, Antonín: Návrh infastruktury železničních uzlů (1. část). *Automa* [online]. 2007, 5, [cit. 2011-6-14].

[3] Bažant, Michael, Kavička, Antonín: Návrh infastruktury železničních uzlů. *Automa* [online]. 2007, 8-9, 102-104 [cit. 2011-6-14].

[4] Billionnet, A.: Using integer programming to solve the train-platforming problem. *Transportation Science, 2003, 37 (2), p. 213-222.*

[5] Carey, M. and Carville, S.: Scheduling and platforming trains at busy komplex stations. Transportation Research Part A: Policy and Practice*, 2003, 37 (3), p. 195-224.*

[6] Corman, F., D'Ariano, A., Pranzo, M., Hansen, I.A.: Effectiveness of dynamic reordering and rerouting of trains in a complicated and densely occupied station area. *3rd International Seminar on Railway Operations Modelling and Analysis, Zürich, 2009.*

[7] Chakroborty, P., Vikram, D.: Optimum assignment of trains to platforms under partial schedule compliance. *Transportation Research Part B: Methodological*, 2008, 42(2), p. 169–184.

[8] Jánošíková, Ľudmila, Bažant, Michael, Kavička, Antonín: Podpora optimálního operativního plánování provozu v osobních železničních stanicích. *Perner's Contacts*. 2009, 4(III.), 97-114.

[9] Kroon, Leo G., Romeijn, Edwin H., Zwaneveld, Peter J.: Routing trains through railway stations: complexity issues. *European Journal of Operational Research*. 1997(98), 485-498.

[10] Záhorová, Věra: *Vliv struktury báze pravidel fuzzy regulátoru na kvalitu rozhodování*. *Perner's Contacts*. 2008, 3(5), 300-307.

[11] Zwaneveld, P.J., Kroon, L.G., Van Hoesel, S.P.M.: Routing trains through a railway station based on a node packing model. *European Journal of Operational Research, 2001, 128(1), p. 14–33.*

Backtesting of market risk estimation assuming various copula functions

Aleš Kresta[1]

Abstract. Market risk estimation is a challenging and no less important task of all financial institutions, which requires the modeling of portfolio returns. When modeling the portfolio returns, we are interested in modeling both the distributions of individual assets returns and dependency of these marginal distributions. The useful tool for dependency modeling are copula functions. The task of this article is to compare the utilization of various copula functions, specifically Gaussian, Student and some other types of copula functions, for portfolio returns modeling and subsequent VaR estimation. As marginal distributions normal inverse Gaussian model (NIG) and also normal distribution are assumed in the paper. These two marginal distributions both joined by chosen copula functions are backtested on time series of historical returns of portfolios dependent on both stock market indices and foreign exchange rates.

Keywords: backtesting, market risk, model validation, subordinated Lévy model, copula function.

JEL Classification: G15, G21, G22
AMS Classification: 60G51

1 Introduction

Financial risk management is an important part of all financial institutions such as banks and insurance companies. In order to manage the risk well we have to be able to measure the risk soundly. Since the Gaussian model is not suitable for returns modeling, some alternative models for market risk modeling have been tested recently. Rank [13] analyzed various marginal distributions coupled together by copula functions for risk estimation, Alexander and Sheedy [1] assumed Gaussian/Student/GARCH/Empirical models for a simple positions. Also Lévy models are suitable for marginal risk modeling as showed Tichý [15], who assumed a variance gamma model (VG model) and a normal inverse Gaussian model (NIG model) coupled together by elliptical copula functions. As the author showed the VG model and the NIG model provided almost the same results. Hence Kresta and Tichý [9; 10] assumed only the NIG model as it is computationally less costly to evaluate its inverse distribution function which is needed in copula modeling framework.

In this paper we extend the previous researches also on Archimedean copula functions. The goal of the paper is to backtest various elliptical and Archimedean copula functions with marginal distributions in form of the Gaussian normal distribution and the NIG model and to compare the quantity of observed exceptions obtained utilizing this models.

We proceed as follows. First, the normal inverse Gaussian model is characterized. Then the copula functions are defined with the focus on the most important elliptical and Archimedean copula functions. Afterwards the backtesting procedure is introduced and statistical test concerning the results of the backtesting procedure is described. In empirical part of the paper we present the results obtained by utilization of the Gaussian normal distribution and the NIG model coupled together by different copula functions for Value at Risk (VaR) estimation of chosen portfolios.

2 Methodology

In this paper we utilize the copula functions for market risk estimation in terms of VaR. Hence, we assume several distinct risk factors, i.e. a marginal (independent) distributions, for which the NIG model (subsection 2.1) is utilized. These marginal distributions are grouped together by various copula functions (subsection 2.2) and market risk models are formed. These models are backtested on historical data of chosen portfolios of FX rates and equity indices. The backtesting procedure is described in subsection 2.3.

[1] VŠB – Technical University of Ostrava/Faculty of Economics, Department of Finance, Sokolská tř. 33, Ostrava, ales.kresta@seznam.cz.

2.1 Normal inverse Gaussian model

The normal inverse Gaussian process belongs to family of Lévy models. The most recent and complete monographs on the theory and application of Lévy models are e.g. [2; 5].

Generally, a Lévy process is a stochastic process, which is zero at origin, its path in time is right-continuous with left limits and its main property is that it is of independent and stationary increments. Another common feature is a so called stochastic continuity. Moreover, the related probability distribution of increments must be infinitely divisible. Concerning the probability distribution the crucial theorem is the Lévy-Khintchine formula:

$$\Phi(u) = i\gamma u - \frac{1}{2}\sigma^2 u^2 + \int_{-\infty}^{\infty} \left(\exp(iux) - 1 - iuxI_{|x|<1}\right)\nu(dx). \tag{1}$$

For a given infinitely divisible distribution, we can define the triplet of Lévy characteristics,

$$\left\{\gamma, \sigma^2, \nu(dx)\right\}.$$

The former two define the drift of the process (deterministic part) and its diffusion. The latter is a Lévy measure. If it can be formulated as $\nu(dx) = u(x)dx$, it is a Lévy density.

Let X be a Brownian motion. If we substitute a standard time t in Brownian motion X,

$$X(t;\mu;\sigma) = \mu t + \sigma Z(t), \tag{2}$$

by a suitable function $l(t)$ as follows,

$$X(l(t);\theta;\vartheta) = \theta l(t) + \vartheta Z(l(t)) = \theta l(t) + \vartheta\varepsilon\sqrt{l(t)}, \tag{3}$$

we get a subordinated Lévy model. Due to the simplicity (tempered stable subordinators with known density functions in the closed form), the most suitable candidates for function $l(t)$ seem to be either (i) the VG model – the overall process is driven by a gamma process from the gamma distribution or (ii) the NIG model – the subordinator is given by an inverse Gaussian process based on the inverse Gaussian distribution.

In financial literature the NIG model was introduced by Barndorff-Nielsen [3]. Assuming parameters $\alpha > 0$, $-\alpha < \beta < \alpha$ and $\delta > 0$, the NIG model is defined by its characteristic function,

$$\phi_{NIG}(u,\alpha,\beta,\delta) = \exp\left(\delta\sqrt{\alpha^2 - \beta^2} - \delta\sqrt{\alpha^2 - (\beta + iu)^2}\right), \tag{4}$$

or it can be viewed as a subordinated Lévy model by assuming inverse Gaussian process $I(t;\nu)$ as $l(t)$ in (3),

$$X(I(t;\nu);\theta,\vartheta) = \theta I(t;\nu) + \vartheta\varepsilon\sqrt{I(t;\nu)}. \tag{5}$$

The parameters θ, ϑ and ν can be calculated from parameters α, β and δ (and vice versa) as follows,

$$\theta = \frac{\delta\beta}{\sqrt{\alpha^2 - \beta^2}},\ \vartheta = \frac{\sqrt{\delta\sqrt{\alpha^2 - \beta^2}}}{\sqrt{\alpha^2 - \beta^2}},\ \nu = \frac{1}{\delta\sqrt{\alpha^2 - \beta^2}}. \tag{6}$$

2.2 Copula functions

A useful tool for dependency modeling are the copula functions, i.e. the projection of the dependency among particular distribution functions into $[0,1]$,

$$C : [0,1]^n \to [0,1] \text{ on } R^n,\ n \in \{2,3,...\}. \tag{7}$$

Basic reference for the theory of copula functions can be found in [12], while [8; 13] target mainly on the application issues in finance. Alternatively, Lévy processes can be coupled on the basis of Lévy measures by Lévy copula functions. However, this approach is not necessary in our case.

Actually, any copula function can be regarded as a multidimensional distribution function with marginals in the form of standardized uniform distribution.

For simplicity, assume two potentially dependent random variables with marginal distribution functions F_X, F_Y and joint distribution function $F_{X,Y}$. Then, following the Sklar's theorem [14]:

$$F_{X,Y}(x,y) = C(F_X(x), F_Y(y)). \tag{8}$$

If both F_X and F_Y are continuous, a copula function C is unique. Sklar's theorem implies also an inverse relation,

$$C(u,v) = F_{X,Y}\left(F_X^{-1}(u), F_Y^{-1}(v)\right). \tag{9}$$

The formulation above should be understood such that the joint distribution function gives us two distinct information: (i) marginal distributions of random variables, (ii) dependency function of the distributions. Hence, while the former is given by $F_X(x)$ and $F_Y(y)$, copula function specifies the dependency. Only when we put both information together, we get sufficient knowledge about the pair of random variables X, Y.

Elliptical and Archimedean copula functions

With some simplicity we can distinguish the elliptical copula functions and Archimedean copula functions. The elliptical copula functions are based on some elliptical joint distribution, such as Gaussian copula function,

$$C_R^{Ga}(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-R^2}} e^{\left[\frac{2Rst-s^2-t^2}{2-2R^2}\right]} ds\, dt \tag{10}$$

where R is the correlation coefficient, or Student copula function based on the Student t distribution,

$$C_{R,v}^{St}(u,v) = \int_{-\infty}^{t_v^{-1}(u)} \int_{-\infty}^{t_v^{-1}(v)} \frac{1}{2\pi\sqrt{1-R^2}} \left[1 + \frac{s^2+t^2-2Rst}{v\left(1-R^2\right)}\right]^{-\frac{v+2}{2}} ds\, dt, \tag{11}$$

where again R is the correlation coefficient and v stands for degrees of freedom of the Student t distribution.

On the other hand, Archimedean copula functions are defined on the basis of function ϕ called generator. Generator is continuous, decreasing and convex function such that $\phi(1) = 0$ and for a strict generator also stands that $\phi(0) = +\infty$. Archimedean copula functions can be defined then as follows,

$$C_{\phi,\phi^{[-1]}}^{Arch}(u,v) = \phi^{[-1]}(\phi(u), \phi(v)), \tag{12}$$

where $\phi^{[-1]}$ is the pseudo-inverse function such that $\phi^{[-1]}(\phi(v)) = v$ for every $v \in \langle 0;1 \rangle$. The most known Archimedean copula functions are: Gumbel copula function [7],

$$C_a^{Gl}(u,v) = \exp\left\{-\left[(-\ln u)^a + (-\ln v)^a\right]^{\frac{1}{a}}\right\}, \tag{13}$$

Clayton copula function [4],

$$C_a^{Cl}(u,v) = \max\left[\left(u^{-a} + v^{-a} - 1\right)^{-\frac{1}{a}}, 0\right], \tag{14}$$

and Frank copula function [6],

$$C_a^{Fr}(u,v) = -\frac{1}{a}\ln\left[1 + \frac{\left(e^{-au}-1\right)\left(e^{-av}-1\right)}{e^{-a}-1}\right]. \tag{15}$$

Parameters estimation

There exist three main approaches to parameter estimation for copula function based dependency modeling: exact maximum likelihood method (EMLM), inference function for margins (IFM), and canonical maximum likelihood (CML). While for the former all parameters are estimated within one step, which might be very time consuming (mainly for high dimensional problems or complicated marginal distributions), the latter two methods

are based on the estimation of the parameters for the marginal distribution and parameters for the copula function separately. While assuming IFM, marginal distributions are estimated in the first step and the copula function in the second one, for CML instead of parametric margins empirical distributions are used.

2.3 Backtesting

Within the backtesting procedure, the ability of a given model to estimate the future losses is tested. Loosely speaking, backtesting is based on the estimation of the risk (mostly measured as VaR) at time t for time $t + \Delta t$, where Δt is usually (in line with the standards for bank supervision defined within Basel II) set to one business day, and compared with the true loss observed at time $t + \Delta t$. This procedure is applied for moving time window over the whole utilized data set.

Within the backtesting procedure on a given time series two situations can arise – the loss is higher or lower than its estimation (from the stochastic point of view, the equality shouldn't arise). While the former case is denoted by 1 as an exception, the latter one is denoted by zero. In this way, we get the sequence of logical values corresponding to the fact whether the exception occurred or not. On this sequence, it can be tested whether the number of ones (exceptions) corresponds with the assumption, i.e. $\alpha \cdot n$ (where n is the length of the sequence), whether the estimation is valid either unconditionally or conditionally, whether bunching is present, etc.

The quantity of exceptions can be tested by Kupiec's test [11] which is derived from a relative amount of exceptions. Thus it tests whether the number of exceptions is from the statistical point of view different from the assumption. A given likelihood ratio on the basis of χ^2 probability distribution with one degree of freedom is formulated as follows:

$$LR = -2\ln\left[\frac{\pi_{ex}^{n_1}\left(1-\pi_{ex}\right)^{n_0}}{\pi_{obs}^{n_1}\left(1-\pi_{obs}\right)^{n_0}}\right], \tag{16}$$

where π_{ex} is expected probability of exception occurring, π_{obs} is observed probability of exception occurring, $\pi_{obs} = \frac{n_1}{n_0 + n_1}$, n_0 is the number of zeros and n_1 is the number of ones (the quantity of exceptions).

3 Results

The data set we consider in this study comprises of daily closing prices of four well established equity indices – Down Jones Industrial Average (DJI) from the US market, FTSE 100 (FTSE) from London (UK), Nikkei 225 (N225) from Tokyo (Japan) and Swiss Market Index (SMI) from Switzerland – over preceding 20 years (January 1, 1991 to August 31, 2011). However, the indices are denominated in four distinct currencies, in particular the US dollar (USD), British pound (GBP), Japan yen (JPY) and Swiss franc (CHF). This fact extends our data set to 8 distinct time series. For all currencies we assume the foreign exchange rate to euro (EUR). Since the trading days at particular markets are not always harmonized, we have to interpolate the missing data. In this way we get eight time series of 5,376 log-returns.

Basic descriptive statistics of daily log-returns are apparent from Table 1. In particular for each index and FX rate the minimal and maximum return, its mean (expected value), median, standard deviation and two higher moments, the skewness and kurtosis, are recorded.

Characteristic	DJI	FTSE	N225	SSMI	USD	GBP	JPY	CHF
minimum	-8.20%	-9.26%	-12.11%	-8.38%	-4.06%	-3.89%	-3.90%	-4.58%
maximum	10.51%	9.38%	10.09%	10.79%	4.82%	2.83%	5.93%	3.26%
mean	0.03%	0.02%	-0.02%	0.03%	0.00%	0.00%	0.01%	0.01%
median	0.05%	0.04%	0.00%	0.07%	-0.01%	0.00%	-0.03%	0.00%
standard deviation	1.10%	1.13%	1.47%	1.17%	0.65%	0.48%	0.76%	0.36%
skewness	-0.105	-0.111	-0.323	-0.128	0.142	-0.421	0.415	-0.098
kurtosis	11.740	9.644	8.283	9.323	5.808	7.863	6.937	17.593

Table 1 Basic descriptive statistics of daily log-returns of utilized dataset

From these assets we construct two types of simple portfolios: (i) we assume the equity index and the corresponding FX rate, thus the investment into the index from euro investor point of view, and (ii) the portfolios of USD/EUR currency pair and one from the other three currency pairs. The difference in these two types of portfolios is in the sign of the correlation coefficient. While the indices and corresponding currencies are correlated negatively, the correlations between the FX rates are positive.

For modeling purpose we assume two types of marginal distributions: (i) normal distribution for its simplicity and (ii) normal inverse Gaussian distribution for its ability to model the higher moments. According to backtesting procedure for each of 3,376 days (first 2,000 days was left for initial parameters estimation) we estimate the parameters of the models (both marginal distributions and copula functions), simulate 50,000 random returns for each asset and then estimate the portfolio VaR for chosen day. By comparison of estimated VaR with true losses we get the sequence of logical values indicating, whether the exception occurred.

The sums of exceptions for models with normal distribution and Gaussian/Student/Clayton/Gumbel/Frank copula function are summarized in Table 2. The numbers which can be statistically accepted on 10% probability level are denoted in bold. In italics we denoted the numbers of exceptions which are closest to the assumption, we call this cases as the winning. We can see, that the normal distribution is truly not the good model for marginal distributions. For $\alpha = 15\%$ the number of observed exceptions is low – the model overestimate the risk. For $\alpha = 1\%$ and $\alpha = 0.5\%$ the number of exceptions is too high to be statistically accepted. These probability levels are assumed in financial sector for risk estimation, thus normal distribution even coupled together with any copula can not be utilized for risk estimation in financial sector. On the other hand for probability level $\alpha = 5\%$ the normal distribution works good, while almost for all chosen portfolios the number of exceptions can be statistically accepted on 10% probability level. The best results are obtained when Gaussian or Frank copula function is utilized. As normal distribution and Gaussian copula function is nothing more than joint normal distribution, we can conclude, that for estimation of VaR (only) at 5% probability level the joint normal distribution is sufficient.

Portfolio	$\alpha = 0.005$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.15$
Assumption	16,88	33,76	168,8	506,4
DJI & USD	47/*41*/42/44/46	63/60/*55*/56/60	**173/174/162/160/*172***	471/***477***/452/450/471
FTSE & GBP	57/56/*50*/52/54	80/78/*74*/75/79	**190**/191/***181*/186/190**	456/*464*/449/448/453
N225 & JPY	38/35/*28*/31/34	53/50/***40*/43**/52	**162/164**/127/131/***165***	437/443/375/379/*447*
SSMI & CHF	56/55/46/*45*/50	82/81/*71*/*71*/79	**186/184/158/*161*/185**	450/449/403/403/*455*
USD & GBP	39/36/*29*/50/53	57/55/*45*/67/69	***169*/168/158/179/174**	462/469/*470*/*470*/442
USD & JPY	29/**23**/***19***/36/40	45/**43**/***35***/53/63	***168*/*168*/161/174/167**	439/451/*457*/447/419
USD & CHF	46/*36*/38/41/48	69/*64*/65/69/74	**187**/194/***168*/181/190**	456/***475***/439/435/460
Sum of significant cases	0/1/1/0/0	0/1/2/1/0	7/5/6/6/7	0/2/0/0/0
Sum of winning cases	0/2/4/1/0	0/1/6/1/0	2/1/2/1/2	0/3/2/1/2

Table 2 The numbers of exceptions for normal distribution and Gaussian/Student/Clayton/Gumbel/Frank copula functions. The closest numbers to the assumption is denoted in italics and on 10% significance level statistically significant numbers of exceptions are denoted in bold.

In Table 3 the quantity of exceptions for NIG model and various copula functions are summarized. Again we denoted the statistically acceptable numbers of exceptions in bold and the winning numbers of exceptions in italics. Compared to the normal distribution there is generally more cases for which the quantity of exceptions can be statistically accepted. For important probability levels 1% and 0.5% the NIG model coupled together with the Clayton copula can be accepted for almost all portfolios (except portfolio USD & CHF). Also utilizing Student copula function the results are very good (the model is insufficient for FTSE & GBP portfolio and again USD & CHF). Other copulas show worse results, but the numbers of exceptions are still half the numbers when normal distribution is utilized. Also for $\alpha = 5\%$ the results of Clayton copula function are very good – only for N225 & JPY the number of observer exceptions is, surprisingly, too low to be statistically accepted. On the other hand, for $\alpha = 15\%$ the results are poor – only for 4 out of 7 portfolios the results can be accepted for Gaussian/Gumbel/Frank copulas and results of Student/Clayton copulas are even worse.

From the results presented in Table 3 we can conclude that the most appropriate copula for VaR estimation is the Clayton copula function. Except the VaR at probability level $\alpha = 15\%$ the Clayton copula is the best from all chosen copulas in terms of the statistically accepted numbers of exceptions and the winning cases (i.e. the closest numbers of observed exceptions to the assumed numbers). Also Student copula function shows good results.

Portfolio	$\alpha = 0.005$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.15$
Assumption	16,88	33,76	168,8	506,4
DJI & USD	**24**/**23**/***21***/**22**/**22**	***35***/**36**/**36**/**37**/**36**	**190**/196/***172***/**179**/193	560/567/544/*542*/561
FTSE & GBP	25/**23**/***20***/**22**/25	45/44/***43***/46/44	193/194/***184***/**190**/191	559/566/*545*/548/561
N225 & JPY	**20**/**18**/***17***/**20**/**20**	***33***/***33***/**28**/**31**/***33***	***162***/**160**/126/136/**158**	***517***/**525**/444/445/**521**
SSMI & CHF	***17***/***17***/**14**/**16**/***17***	**41**/**42**/***33***/**35**/**36**	206/197/**172**/***169***/194	575/584/***535***/**536**/580
USD & GBP	**18**/***17***/**11**/**21**/26	**38**/***34***/**26**/46/47	**173**/**174**/***165***/**183**/**175**	***515***/**535**/548/**521**/**495**
USD & JPY	**24**/**22**/***18***/30/31	**42**/**38**/***33***/47/50	**187**/193/***182***/198/191	**527**/543/553/**535**/***499***
USD & CHF	27/***22***/**24**/***22***/29	52/47/*44*/50/54	200/207/***187***/192/207	**520**/544/***509***/**499**/**518**
Sum of significant cases	5/7/7/6/3	5/5/6/3/3	4/2/6/4/2	4/2/2/4/4
Sum of winning cases	1/3/4/1/1	2/2/4/0/1	1/0/5/1/0	2/0/3/1/1

Table 3 The numbers of exceptions for NIG model and Gaussian/Student/Clayton/Gumbel/Frank copula functions. The closest numbers to the assumption is denoted in italics and on 10% significance level statistically significant numbers of exceptions are denoted in bold.

4 Conclusion

Unexpectedly high decreases in the prices of financial assets are very challenging task for any risk model. In this paper we compared the risk models composed of NIG model (or Gaussian normal distribution) and various copula functions on the basis of the quantity of exceptions observed. From the provided results it is apparent that the most accurate risk estimations in terms of VaR at probability levels 0.5% and 1% are provided by NIG model and Clayton copula function. Also Student copula function provides good results. We also concluded, that for the estimation of VaR at 5% probability level the joint Gaussian normal distribution is sufficient.

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 and furthermore supported by SP2012/2, a SGS research project of VSB-TU Ostrava.

References

[1] Alexander, C., and Sheedy, E: Developing a stress testing framework based on market risk models. *Journal of Banking and Finance* **32** (2008), 2220-2236.
[2] Applebaum, D.: *Lévy Processes and Stochastic Calculus.* Cambridge University Press, Cambridge, 2004.
[3] Barndorff-Nielsen, O. E.: *Normal inverse Gaussian distributions and the modeling of stock returns.* Aarhus University, Aarhus, 1995. Doctoral dissertation.
[4] Clayton, D. G.: A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65** (1978), 141-151.
[5] Cont, R., and Tankov, P.: *Financial Modelling with Jump Processes.* Chapman & Hall/CRC, London, 2003.
[6] Frank, M. J.: On the simultaneous associativity of F(x, y) and x+y-F(x, y). *Aequationes Mathematicae* **19** (1979), 194-226.
[7] Gumbel, E. J.: Bivariate exponential distributions. *Journal of the American Statistical Association* **55** (1960), 698-707.
[8] Cherubini, G., Luciano, E., and Vecchiato, W.: *Copula Methods in Finance*. Wiley, Chichester, 2004.
[9] Kresta, A., and Tichý, T.: International Equity Risk Modeling by NIG Model. In: *Mathematical methods in economics* (Dlouhý, M., and Skočdopolová, V., eds.). Professional Publishing, Prague, 2011, 401-406.
[10] Kresta, A., and Tichý, T.: International Equity Portfolio Risk Modeling: The case of NIG model and ordinary copula functions. *Finance a úvěr – Czech Journal of Economics and Finance* **61** (2012), 141–151.
[11] Kupiec, P.: Techniques for verifying the accuracy of risk measurement models. *Journal of Derivative* **3** (1995), 73-84.
[12] Nelsen, R. B.: *An Introduction to Copulas*. 2nd ed. Springer, New York, 2006.
[13] Rank, J.: *Copulas: From Theory to Application in Finance.* Risk Books, London, 2006.
[14] Sklar, A.: Fonctions de repartition à n dimensions et leurs marges. *Publications de l'Institut de statistique de l'Université de Paris* **8** (1959), 229-231.
[15] Tichý, T.: Posouzení odhadu měnového rizika portfolia pomocí Lévyho modelů. *Politická ekonomie* **58** (2010), 504-521.

Backtesting of market risk estimation assuming various copula functions

Aleš Kresta[1]

Abstract. Market risk estimation is a challenging and no less important task of all financial institutions, which requires the modeling of portfolio returns. When modeling the portfolio returns, we are interested in modeling both the distributions of individual assets returns and dependency of these marginal distributions. The useful tool for dependency modeling are copula functions. The task of this article is to compare the utilization of various copula functions, specifically Gaussian, Student and some other types of copula functions, for portfolio returns modeling and subsequent VaR estimation. As marginal distributions normal inverse Gaussian model (NIG) and also normal distribution are assumed in the paper. These two marginal distributions both joined by chosen copula functions are backtested on time series of historical returns of portfolios dependent on both stock market indices and foreign exchange rates.

Keywords: backtesting, market risk, model validation, subordinated Lévy model, copula function.

JEL Classification: G15, G21, G22
AMS Classification: 60G51

1 Introduction

Financial risk management is an important part of all financial institutions such as banks and insurance companies. In order to manage the risk well we have to be able to measure the risk soundly. Since the Gaussian model is not suitable for returns modeling, some alternative models for market risk modeling have been tested recently. Rank [13] analyzed various marginal distributions coupled together by copula functions for risk estimation, Alexander and Sheedy [1] assumed Gaussian/Student/GARCH/Empirical models for a simple positions. Also Lévy models are suitable for marginal risk modeling as showed Tichý [15], who assumed a variance gamma model (VG model) and a normal inverse Gaussian model (NIG model) coupled together by elliptical copula functions. As the author showed the VG model and the NIG model provided almost the same results. Hence Kresta and Tichý [9; 10] assumed only the NIG model as it is computationally less costly to evaluate its inverse distribution function which is needed in copula modeling framework.

In this paper we extend the previous researches also on Archimedean copula functions. The goal of the paper is to backtest various elliptical and Archimedean copula functions with marginal distributions in form of the Gaussian normal distribution and the NIG model and to compare the quantity of observed exceptions obtained utilizing this models.

We proceed as follows. First, the normal inverse Gaussian model is characterized. Then the copula functions are defined with the focus on the most important elliptical and Archimedean copula functions. Afterwards the backtesting procedure is introduced and statistical test concerning the results of the backtesting procedure is described. In empirical part of the paper we present the results obtained by utilization of the Gaussian normal distribution and the NIG model coupled together by different copula functions for Value at Risk (VaR) estimation of chosen portfolios.

2 Methodology

In this paper we utilize the copula functions for market risk estimation in terms of VaR. Hence, we assume several distinct risk factors, i.e. a marginal (independent) distributions, for which the NIG model (subsection 2.1) is utilized. These marginal distributions are grouped together by various copula functions (subsection 2.2) and market risk models are formed. These models are backtested on historical data of chosen portfolios of FX rates and equity indices. The backtesting procedure is described in subsection 2.3.

[1] VŠB – Technical University of Ostrava/Faculty of Economics, Department of Finance, Sokolská tř. 33, Ostrava, ales.kresta@seznam.cz.

2.1 Normal inverse Gaussian model

The normal inverse Gaussian process belongs to family of Lévy models. The most recent and complete monographs on the theory and application of Lévy models are e.g. [2; 5].

Generally, a Lévy process is a stochastic process, which is zero at origin, its path in time is right-continuous with left limits and its main property is that it is of independent and stationary increments. Another common feature is a so called stochastic continuity. Moreover, the related probability distribution of increments must be infinitely divisible. Concerning the probability distribution the crucial theorem is the Lévy-Khintchine formula:

$$\Phi(u) = i\gamma u - \frac{1}{2}\sigma^2 u^2 + \int_{-\infty}^{\infty} \left(\exp(iux) - 1 - iuxI_{|x|<1}\right)\nu(dx). \tag{1}$$

For a given infinitely divisible distribution, we can define the triplet of Lévy characteristics,

$$\left\{\gamma, \sigma^2, \nu(dx)\right\}.$$

The former two define the drift of the process (deterministic part) and its diffusion. The latter is a Lévy measure. If it can be formulated as $\nu(dx) = u(x)dx$, it is a Lévy density.

Let X be a Brownian motion. If we substitute a standard time t in Brownian motion X,

$$X(t;\mu;\sigma) = \mu t + \sigma Z(t), \tag{2}$$

by a suitable function $l(t)$ as follows,

$$X(l(t);\theta;\vartheta) = \theta l(t) + \vartheta Z(l(t)) = \theta l(t) + \vartheta\varepsilon\sqrt{l(t)}, \tag{3}$$

we get a subordinated Lévy model. Due to the simplicity (tempered stable subordinators with known density functions in the closed form), the most suitable candidates for function $l(t)$ seem to be either (i) the VG model – the overall process is driven by a gamma process from the gamma distribution or (ii) the NIG model – the subordinator is given by an inverse Gaussian process based on the inverse Gaussian distribution.

In financial literature the NIG model was introduced by Barndorff-Nielsen [3]. Assuming parameters $\alpha > 0$, $-\alpha < \beta < \alpha$ and $\delta > 0$, the NIG model is defined by its characteristic function,

$$\phi_{NIG}(u,\alpha,\beta,\delta) = \exp\left(\delta\sqrt{\alpha^2 - \beta^2} - \delta\sqrt{\alpha^2 - (\beta + iu)^2}\right), \tag{4}$$

or it can be viewed as a subordinated Lévy model by assuming inverse Gaussian process $I(t;\nu)$ as $l(t)$ in (3),

$$X(I(t;\nu);\theta,\vartheta) = \theta I(t;\nu) + \vartheta\varepsilon\sqrt{I(t;\nu)}. \tag{5}$$

The parameters θ, ϑ and ν can be calculated from parameters α, β and δ (and vice versa) as follows,

$$\theta = \frac{\delta\beta}{\sqrt{\alpha^2 - \beta^2}},\ \vartheta = \frac{\sqrt{\delta\sqrt{\alpha^2 - \beta^2}}}{\sqrt{\alpha^2 - \beta^2}},\ \nu = \frac{1}{\delta\sqrt{\alpha^2 - \beta^2}}. \tag{6}$$

2.2 Copula functions

A useful tool for dependency modeling are the copula functions, i.e. the projection of the dependency among particular distribution functions into $[0,1]$,

$$C : [0,1]^n \to [0,1] \text{ on } R^n,\ n \in \{2,3,...\}. \tag{7}$$

Basic reference for the theory of copula functions can be found in [12], while [8; 13] target mainly on the application issues in finance. Alternatively, Lévy processes can be coupled on the basis of Lévy measures by Lévy copula functions. However, this approach is not necessary in our case.

Actually, any copula function can be regarded as a multidimensional distribution function with marginals in the form of standardized uniform distribution.

For simplicity, assume two potentially dependent random variables with marginal distribution functions F_X, F_Y and joint distribution function $F_{X,Y}$. Then, following the Sklar's theorem [14]:

$$F_{X,Y}(x,y) = C(F_X(x), F_Y(y)). \tag{8}$$

If both F_X and F_Y are continuous, a copula function C is unique. Sklar's theorem implies also an inverse relation,

$$C(u,v) = F_{X,Y}\left(F_X^{-1}(u), F_Y^{-1}(v)\right). \tag{9}$$

The formulation above should be understood such that the joint distribution function gives us two distinct information: (i) marginal distributions of random variables, (ii) dependency function of the distributions. Hence, while the former is given by $F_X(x)$ and $F_Y(y)$, copula function specifies the dependency. Only when we put both information together, we get sufficient knowledge about the pair of random variables X, Y.

Elliptical and Archimedean copula functions

With some simplicity we can distinguish the elliptical copula functions and Archimedean copula functions. The elliptical copula functions are based on some elliptical joint distribution, such as Gaussian copula function,

$$C_R^{Ga}(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-R^2}} e^{\left[\frac{2Rst-s^2-t^2}{2-2R^2}\right]} ds\, dt \tag{10}$$

where R is the correlation coefficient, or Student copula function based on the Student t distribution,

$$C_{R,v}^{St}(u,v) = \int_{-\infty}^{t_v^{-1}(u)} \int_{-\infty}^{t_v^{-1}(v)} \frac{1}{2\pi\sqrt{1-R^2}} \left[1 + \frac{s^2+t^2-2Rst}{v\left(1-R^2\right)}\right]^{-\frac{v+2}{2}} ds\, dt, \tag{11}$$

where again R is the correlation coefficient and v stands for degrees of freedom of the Student t distribution.

On the other hand, Archimedean copula functions are defined on the basis of function ϕ called generator. Generator is continuous, decreasing and convex function such that $\phi(1) = 0$ and for a strict generator also stands that $\phi(0) = +\infty$. Archimedean copula functions can be defined then as follows,

$$C_{\phi,\phi^{[-1]}}^{Arch}(u,v) = \phi^{[-1]}(\phi(u), \phi(v)), \tag{12}$$

where $\phi^{[-1]}$ is the pseudo-inverse function such that $\phi^{[-1]}(\phi(v)) = v$ for every $v \in \langle 0;1 \rangle$. The most known Archimedean copula functions are: Gumbel copula function [7],

$$C_a^{Gl}(u,v) = \exp\left\{-\left[(-\ln u)^a + (-\ln v)^a\right]^{\frac{1}{a}}\right\}, \tag{13}$$

Clayton copula function [4],

$$C_a^{Cl}(u,v) = \max\left[\left(u^{-a} + v^{-a} - 1\right)^{-\frac{1}{a}}, 0\right], \tag{14}$$

and Frank copula function [6],

$$C_a^{Fr}(u,v) = -\frac{1}{a}\ln\left[1 + \frac{\left(e^{-au}-1\right)\left(e^{-av}-1\right)}{e^{-a}-1}\right]. \tag{15}$$

Parameters estimation

There exist three main approaches to parameter estimation for copula function based dependency modeling: exact maximum likelihood method (EMLM), inference function for margins (IFM), and canonical maximum likelihood (CML). While for the former all parameters are estimated within one step, which might be very time consuming (mainly for high dimensional problems or complicated marginal distributions), the latter two methods

are based on the estimation of the parameters for the marginal distribution and parameters for the copula function separately. While assuming IFM, marginal distributions are estimated in the first step and the copula function in the second one, for CML instead of parametric margins empirical distributions are used.

2.3 Backtesting

Within the backtesting procedure, the ability of a given model to estimate the future losses is tested. Loosely speaking, backtesting is based on the estimation of the risk (mostly measured as VaR) at time t for time $t + \Delta t$, where Δt is usually (in line with the standards for bank supervision defined within Basel II) set to one business day, and compared with the true loss observed at time $t + \Delta t$. This procedure is applied for moving time window over the whole utilized data set.

Within the backtesting procedure on a given time series two situations can arise – the loss is higher or lower than its estimation (from the stochastic point of view, the equality shouldn't arise). While the former case is denoted by 1 as an exception, the latter one is denoted by zero. In this way, we get the sequence of logical values corresponding to the fact whether the exception occurred or not. On this sequence, it can be tested whether the number of ones (exceptions) corresponds with the assumption, i.e. $\alpha \cdot n$ (where n is the length of the sequence), whether the estimation is valid either unconditionally or conditionally, whether bunching is present, etc.

The quantity of exceptions can be tested by Kupiec's test [11] which is derived from a relative amount of exceptions. Thus it tests whether the number of exceptions is from the statistical point of view different from the assumption. A given likelihood ratio on the basis of χ^2 probability distribution with one degree of freedom is formulated as follows:

$$LR = -2\ln\left[\frac{\pi_{ex}^{n_1}\left(1-\pi_{ex}\right)^{n_0}}{\pi_{obs}^{n_1}\left(1-\pi_{obs}\right)^{n_0}}\right], \tag{16}$$

where π_{ex} is expected probability of exception occurring, π_{obs} is observed probability of exception occurring, $\pi_{obs} = \frac{n_1}{n_0 + n_1}$, n_0 is the number of zeros and n_1 is the number of ones (the quantity of exceptions).

3 Results

The data set we consider in this study comprises of daily closing prices of four well established equity indices – Down Jones Industrial Average (DJI) from the US market, FTSE 100 (FTSE) from London (UK), Nikkei 225 (N225) from Tokyo (Japan) and Swiss Market Index (SMI) from Switzerland – over preceding 20 years (January 1, 1991 to August 31, 2011). However, the indices are denominated in four distinct currencies, in particular the US dollar (USD), British pound (GBP), Japan yen (JPY) and Swiss franc (CHF). This fact extends our data set to 8 distinct time series. For all currencies we assume the foreign exchange rate to euro (EUR). Since the trading days at particular markets are not always harmonized, we have to interpolate the missing data. In this way we get eight time series of 5,376 log-returns.

Basic descriptive statistics of daily log-returns are apparent from Table 1. In particular for each index and FX rate the minimal and maximum return, its mean (expected value), median, standard deviation and two higher moments, the skewness and kurtosis, are recorded.

Characteristic	DJI	FTSE	N225	SSMI	USD	GBP	JPY	CHF
minimum	-8.20%	-9.26%	-12.11%	-8.38%	-4.06%	-3.89%	-3.90%	-4.58%
maximum	10.51%	9.38%	10.09%	10.79%	4.82%	2.83%	5.93%	3.26%
mean	0.03%	0.02%	-0.02%	0.03%	0.00%	0.00%	0.01%	0.01%
median	0.05%	0.04%	0.00%	0.07%	-0.01%	0.00%	-0.03%	0.00%
standard deviation	1.10%	1.13%	1.47%	1.17%	0.65%	0.48%	0.76%	0.36%
skewness	-0.105	-0.111	-0.323	-0.128	0.142	-0.421	0.415	-0.098
kurtosis	11.740	9.644	8.283	9.323	5.808	7.863	6.937	17.593

Table 1 Basic descriptive statistics of daily log-returns of utilized dataset

From these assets we construct two types of simple portfolios: (i) we assume the equity index and the corresponding FX rate, thus the investment into the index from euro investor point of view, and (ii) the portfolios of USD/EUR currency pair and one from the other three currency pairs. The difference in these two types of portfolios is in the sign of the correlation coefficient. While the indices and corresponding currencies are correlated negatively, the correlations between the FX rates are positive.

For modeling purpose we assume two types of marginal distributions: (i) normal distribution for its simplicity and (ii) normal inverse Gaussian distribution for its ability to model the higher moments. According to backtesting procedure for each of 3,376 days (first 2,000 days was left for initial parameters estimation) we estimate the parameters of the models (both marginal distributions and copula functions), simulate 50,000 random returns for each asset and then estimate the portfolio VaR for chosen day. By comparison of estimated VaR with true losses we get the sequence of logical values indicating, whether the exception occurred.

The sums of exceptions for models with normal distribution and Gaussian/Student/Clayton/Gumbel/Frank copula function are summarized in Table 2. The numbers which can be statistically accepted on 10% probability level are denoted in bold. In italics we denoted the numbers of exceptions which are closest to the assumption, we call this cases as the winning. We can see, that the normal distribution is truly not the good model for marginal distributions. For $\alpha = 15\%$ the number of observed exceptions is low – the model overestimate the risk. For $\alpha = 1\%$ and $\alpha = 0.5\%$ the number of exceptions is too high to be statistically accepted. These probability levels are assumed in financial sector for risk estimation, thus normal distribution even coupled together with any copula can not be utilized for risk estimation in financial sector. On the other hand for probability level $\alpha = 5\%$ the normal distribution works good, while almost for all chosen portfolios the number of exceptions can be statistically accepted on 10% probability level. The best results are obtained when Gaussian or Frank copula function is utilized. As normal distribution and Gaussian copula function is nothing more than joint normal distribution, we can conclude, that for estimation of VaR (only) at 5% probability level the joint normal distribution is sufficient.

Portfolio	$\alpha = 0.005$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.15$
Assumption	16,88	33,76	168,8	506,4
DJI & USD	47/*41*/42/44/46	63/60/*55*/56/60	**173/174/162/160/*172***	471/***477***/452/450/471
FTSE & GBP	57/56/*50*/52/54	80/78/*74*/75/79	**190**/191/***181*/186/190**	456/*464*/449/448/453
N225 & JPY	38/35/*28*/31/34	53/50/***40*/43**/52	**162/164**/127/131/***165***	437/443/375/379/*447*
SSMI & CHF	56/55/46/*45*/50	82/81/*71*/*71*/79	**186/184/158/*161*/185**	450/449/403/403/*455*
USD & GBP	39/36/*29*/50/53	57/55/*45*/67/69	***169*/168/158/179/174**	462/469/*470*/*470*/442
USD & JPY	29/**23**/***19***/36/40	45/**43**/***35***/53/63	***168*/*168*/161/174/167**	439/451/*457*/447/419
USD & CHF	46/*36*/38/41/48	69/*64*/65/69/74	**187**/194/***168*/181/190**	456/***475***/439/435/460
Sum of significant cases	0/1/1/0/0	0/1/2/1/0	7/5/6/6/7	0/2/0/0/0
Sum of winning cases	0/2/4/1/0	0/1/6/1/0	2/1/2/1/2	0/3/2/1/2

Table 2 The numbers of exceptions for normal distribution and Gaussian/Student/Clayton/Gumbel/Frank copula functions. The closest numbers to the assumption is denoted in italics and on 10% significance level statistically significant numbers of exceptions are denoted in bold.

In Table 3 the quantity of exceptions for NIG model and various copula functions are summarized. Again we denoted the statistically acceptable numbers of exceptions in bold and the winning numbers of exceptions in italics. Compared to the normal distribution there is generally more cases for which the quantity of exceptions can be statistically accepted. For important probability levels 1% and 0.5% the NIG model coupled together with the Clayton copula can be accepted for almost all portfolios (except portfolio USD & CHF). Also utilizing Student copula function the results are very good (the model is insufficient for FTSE & GBP portfolio and again USD & CHF). Other copulas show worse results, but the numbers of exceptions are still half the numbers when normal distribution is utilized. Also for $\alpha = 5\%$ the results of Clayton copula function are very good – only for N225 & JPY the number of observer exceptions is, surprisingly, too low to be statistically accepted. On the other hand, for $\alpha = 15\%$ the results are poor – only for 4 out of 7 portfolios the results can be accepted for Gaussian/Gumbel/Frank copulas and results of Student/Clayton copulas are even worse.

From the results presented in Table 3 we can conclude that the most appropriate copula for VaR estimation is the Clayton copula function. Except the VaR at probability level $\alpha = 15\%$ the Clayton copula is the best from all chosen copulas in terms of the statistically accepted numbers of exceptions and the winning cases (i.e. the closest numbers of observed exceptions to the assumed numbers). Also Student copula function shows good results.

Portfolio	$\alpha = 0.005$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.15$
Assumption	16,88	33,76	168,8	506,4
DJI & USD	**24**/**23**/***21***/**22**/**22**	***35***/**36**/**36**/**37**/**36**	**190**/196/***172***/**179**/193	560/567/544/*542*/561
FTSE & GBP	25/**23**/***20***/**22**/25	45/44/***43***/46/44	193/194/***184***/**190**/191	559/566/*545*/548/561
N225 & JPY	**20**/**18**/***17***/**20**/**20**	***33***/***33***/**28**/**31**/***33***	***162***/**160**/126/136/**158**	***517***/**525**/444/445/**521**
SSMI & CHF	***17***/***17***/**14**/**16**/***17***	**41**/**42**/***33***/**35**/**36**	206/197/**172**/***169***/194	575/584/***535***/**536**/580
USD & GBP	**18**/***17***/**11**/**21**/26	**38**/***34***/**26**/46/47	**173**/**174**/***165***/**183**/**175**	***515***/**535**/548/**521**/**495**
USD & JPY	**24**/**22**/***18***/30/31	**42**/**38**/***33***/47/50	**187**/193/***182***/198/191	**527**/543/553/**535**/***499***
USD & CHF	27/***22***/**24**/***22***/29	52/47/*44*/50/54	200/207/***187***/192/207	**520**/544/***509***/**499**/**518**
Sum of significant cases	5/7/7/6/3	5/5/6/3/3	4/2/6/4/2	4/2/2/4/4
Sum of winning cases	1/3/4/1/1	2/2/4/0/1	1/0/5/1/0	2/0/3/1/1

Table 3 The numbers of exceptions for NIG model and Gaussian/Student/Clayton/Gumbel/Frank copula functions. The closest numbers to the assumption is denoted in italics and on 10% significance level statistically significant numbers of exceptions are denoted in bold.

4 Conclusion

Unexpectedly high decreases in the prices of financial assets are very challenging task for any risk model. In this paper we compared the risk models composed of NIG model (or Gaussian normal distribution) and various copula functions on the basis of the quantity of exceptions observed. From the provided results it is apparent that the most accurate risk estimations in terms of VaR at probability levels 0.5% and 1% are provided by NIG model and Clayton copula function. Also Student copula function provides good results. We also concluded, that for the estimation of VaR at 5% probability level the joint Gaussian normal distribution is sufficient.

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 and furthermore supported by SP2012/2, a SGS research project of VSB-TU Ostrava.

References

[1] Alexander, C., and Sheedy, E: Developing a stress testing framework based on market risk models. *Journal of Banking and Finance* **32** (2008), 2220-2236.
[2] Applebaum, D.: *Lévy Processes and Stochastic Calculus.* Cambridge University Press, Cambridge, 2004.
[3] Barndorff-Nielsen, O. E.: *Normal inverse Gaussian distributions and the modeling of stock returns.* Aarhus University, Aarhus, 1995. Doctoral dissertation.
[4] Clayton, D. G.: A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65** (1978), 141-151.
[5] Cont, R., and Tankov, P.: *Financial Modelling with Jump Processes.* Chapman & Hall/CRC, London, 2003.
[6] Frank, M. J.: On the simultaneous associativity of F(x, y) and x+y-F(x, y). *Aequationes Mathematicae* **19** (1979), 194-226.
[7] Gumbel, E. J.: Bivariate exponential distributions. *Journal of the American Statistical Association* **55** (1960), 698-707.
[8] Cherubini, G., Luciano, E., and Vecchiato, W.: *Copula Methods in Finance*. Wiley, Chichester, 2004.
[9] Kresta, A., and Tichý, T.: International Equity Risk Modeling by NIG Model. In: *Mathematical methods in economics* (Dlouhý, M., and Skočdopolová, V., eds.). Professional Publishing, Prague, 2011, 401-406.
[10] Kresta, A., and Tichý, T.: International Equity Portfolio Risk Modeling: The case of NIG model and ordinary copula functions. *Finance a úvěr – Czech Journal of Economics and Finance* **61** (2012), 141–151.
[11] Kupiec, P.: Techniques for verifying the accuracy of risk measurement models. *Journal of Derivative* **3** (1995), 73-84.
[12] Nelsen, R. B.: *An Introduction to Copulas*. 2nd ed. Springer, New York, 2006.
[13] Rank, J.: *Copulas: From Theory to Application in Finance.* Risk Books, London, 2006.
[14] Sklar, A.: Fonctions de repartition à n dimensions et leurs marges. *Publications de l'Institut de statistique de l'Université de Paris* **8** (1959), 229-231.
[15] Tichý, T.: Posouzení odhadu měnového rizika portfolia pomocí Lévyho modelů. *Politická ekonomie* **58** (2010), 504-521.

Non-stationary volatility with highly anti-persistent increments: An alternative paradigm in volatility modeling?

Ladislav Krištoufek [1]

Abstract. We introduce the alternative paradigm to volatility modeling. On the example of three stocks of highly capitalized companies, we show that volatility process is non-stationary and its logarithmic transformation together with the logarithmic increments are approximately normally distributed while the latter are strongly anti-persistent. Together with the assertion that logarithmic returns are normally distributed, and uncorrelated with time-varying volatility, we propose the new returns-generating process, which is able to remarkably mimic the real-world series and the standard stylized facts – uncorrelated returns with heavy tails, strongly autocorrelated absolute returns and volatility clustering. The proposed methodology opens a wholly new field in research of financial volatility.

Keywords: volatility, anti-persistence, non-stationarity

JEL classification: C53, C58, G17

1 Introduction

Accurate modeling and forecasting of volatility is one of the biggest challenges in financial economics and econometrics. Historically, there are four major groups of volatility forecasting approaches – historical volatility, conditional heteroskedasticity models, stochastic volatility models, and implied volatility models. These approaches are nicely reviewed and compared in two studies of Poon & Granger [13, 14]. Poon & Granger [14] argue that the implied volatility models outperform the others, followed by the historical volatility models, in volatility forecasting. This is a rather interesting, and disturbing, finding since the Black-Scholes formula is known to be based on highly unrealistic assumptions of the returns process. This might imply that the other approaches are actually not optimal and there is some other approach closer to reality.

In this paper, we propose a new approach to volatility modeling. Based on a simple statistical analysis, we show that volatility can be effectively modeled as a non-stationary process with highly anti-persistent logarithmic increments, which are, moreover, normally distributed. By doing so, we are able to mimic the basic stylized facts of the financial returns – no autocorrelation, highly persistent absolute returns, non-normality, fat tails, and volatility clustering [3]. Analyzing the real-world data, we are able to state seven basic Claims and argue that the logarithmic returns are uncorrelated and normally distributed with approximately zero mean and time-varying standard deviation. The logarithm of the standard deviation is non-stationary and approximately normally distributed with approximately normally distributed increments which are stationary and highly anti-persistent. Based on these findings, we are able to reconstruct the series of returns, which resemble the actual financial returns and the stylized facts very closely.

The paper is structured as follows. In Section 2, we present the basic methodology. In Section 3, we describe the dataset and present the crucial findings about the process of volatility. In Section 4, we show simulations for estimated parameters. Section 5 concludes.

[1]Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, CZ-182 08, Prague, Czech Republic; Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Opletalova 26, CZ-110 00, Prague, Czech Republic, e-mail: kristouf@utia.cas.cz

2 Methodology

Long-range dependence is highly connected to the Hurst effect, i.e. a situation characteristic by long periods when the series is above the mean which are followed by long periods when the series is below the mean of the series while the series still remains stationary, and so also Hurst exponent H. A critical value of Hurst exponent is 0.5 and suggests two possible processes. H being equal to 0.5 implies either an uncorrelated or a short-term dependent process. For $H > 0.5$, auto-covariances of the process are positive at all lags so that the process is labeled as persistent. On the other hand, for $H < 0.5$, the process is said to be long-range dependent with negative correlations or anti-persistent. The persistent process is visibly locally trending while the anti-persistent process switches signs more frequently than a random process would [1].

Hurst exponent H is connected to parameter d of fractional integration so that $H = d + 0.5$. Long-range dependent processes are frequently defined in two domains – time and frequency:

- A stationary process with an autocorrelation function $\rho(k)$ decaying as $\rho(k) \propto k^{2H-2}$ for $k \to \infty$ is called long-range dependent with Hurst exponent H.
- A stationary process with a spectral density $f(\lambda)$ following $f(\lambda) \propto |\lambda|^{1-2H}$ for $\lambda \to 0$ is called long-range dependent with Hurst exponent H.

Note that a notion of long-range dependence is tightly connected to stationarity of the process, which is obvious from both presented definitions. Without stationarity, there is no standard long-range dependence. To check for (non-)stationarity, we apply two standard tests – ADF [5] and KPSS [11]. ADF has a null hypothesis of a unit root and can take lags of the differenced series into consideration and thus controlling for the memory effects. KPSS, on contrary, has a stationarity null hypothesis and also can control for memory effects with a use of auto-covariance adjusted variance with Barlett weights [8].

As we are interested primary in long-range dependence and not necessarily in a specific value of H or d, we use (modified) rescaled range and rescaled variance analyses for testing the presence of long-range dependence in the studied process. For more details, see [12, 8].

Lo [12] proposed the modified rescaled range analysis (M-R/S) as a generalization of the classical rescaled range analysis [10] to control for short-term correlations. With a use of adjusted standard deviation $S^M = \sqrt{S^2 + 2\sum_{j=1}^{\xi} \gamma_j \left(1 - \frac{j}{\xi+1}\right)}$, where S^2 is the standard deviation of the returns, γ_j is the jth auto covariance of the process up to lag ξ, the V statistic is defined as $V_\upsilon = \frac{(R/S)_\upsilon}{\sqrt{\upsilon}}$, where R/S is a range of the profile of the series (cumulative deviations from the mean of the original series) standardized by the standard deviation of the returns process. For our purposes, we set $\upsilon = T$ to test for the presence of long-range dependence in the underlying process by the means of V statistic [12, 10].

Rescaled variance analysis (V/S) was proposed [8] as a modified version of KPSS statistic, which is usually used for testing of stationarity but was also shown to have good power for series with long-term memory. The procedure is very similar to the modified rescaled range analysis and differs in a use of a sample variance of the profile of the series instead of the range. As an alternative to the V statistic, the M statistic is defined as $M_\upsilon = \frac{var(X_{t,I_n})}{\upsilon (S^M_{I_n})^2}$. Note that the modified standard deviation S^M is used so that the method is robust to short-range dependence as well. Variance of the M statistic is much lower than the one of R/S and $M - R/S$ so that the confidence intervals are much narrower and the method is thus more reliable. Similarly to R/S and $M - R/S$, one needs to follow the steps of the R/S if estimating H or set $\upsilon = T$ and construct M statistic if only testing for the presence of long-range dependence in the process, which is actually the case in our study.

The bootstrap method has been proposed to deal with the statistical properties of small samples [6]. The basic notion behind the procedure is resampling with replacement from the original series and repeated estimation of a specific parameter. By shuffling, a distribution of the original series remains unchanged while possible dependencies are distorted. Hypothesis can be then tested with a use of p-values based on the bootstrapped estimates. For our purposes and for the time series analysis in general, the simple bootstrap is not enough as the shuffling rids us not only from the long-range dependence but the short-range dependencies as well. Srinivas & Srinivasan [15] proposed a modified method which retains the short-term dependence characteristics but lacks the long one – the moving block bootstrap with pre-whitening and post-blackening.

For our purposes, we use AR(1) process for pre-whitening and post-blackening and $\zeta = 20$. Such choice of ζ should be sufficient for ridding of the potential long-range dependence while the other properties remain similar to the original process. The procedure is repeated for lags $\xi = 0, 1, 2, 5, 10, 15, 30, 50, 75, 100$ for both modified rescaled range analysis and rescaled variance analysis with $B = 1000$ bootstrap repetitions. As will be visible in the following sections, the analyzed series are on the edge between stationarity and non-stationarity, and we will eventually analyze the first differences of the series. However, for such boundary cases, there is a high risk of over-differencing which would inflict MA(1) process in the series. To control for this, we also apply the moving block bootstrap with the same parameters as noted before but with MA(1) for pre-whitening and post-blackening. This way, we can be more confident about our findings while controlling for the most problematic cases. Let us now turn to the volatility estimate choice.

There are various estimators of daily volatility (or variance) ranging from very simple absolute and squared returns through model based estimators (e.g. GARCH or implied volatility based) to range-based estimators and realized variance [2]. From many possibilities, we choose Garman–Klass estimator [7]:

$$\widehat{\sigma^2_{GK,t}} = \frac{(\log(H_t/L_t))^2}{2} - (2\log 2 - 1)(\log(C_t/O_t))^2 \tag{1}$$

where H_t and L_t are daily highs and lows, respectively, and C_t and O_t are daily closing and opening prices, respectively. This estimator does not take overnight volatility into consideration but is very simple and efficient (much more efficient than absolute and various power-returns, comparable with other range-based measures and less efficient than the realized variance) [2]. Even though the realized variance would be a more efficient choice, it is not easily obtainable for all assets while for the Garman–Klass estimator, all necessary variables are available freely for practically all financial assets.

3 Data and statistical analysis

We analyze series of three stocks with one of the highest capitalizations in the US markets – AAPL (Apple Inc.), IBM (International Business Machines Corporation) and MSFT (Microsoft Corporation) – between 3.1.2000 and 29.2.2012 (3059 observations). Even though all three companies are technological, they underwent very different dynamics during the analyzed period. Apple, as a favorite of the most recent days, has grown remarkably while IBM and Microsoft have been rather stagnant. IBM and MSFT experienced very similar dynamics of volatility as well as its levels. AAPL, on contrary, shows markedly higher average values of volatility with more extreme values than the other two. Nevertheless, we do not observe any drastic jumps in the volatility levels and we rather find smooth transitions from lower to higher levels or vice versa.

As the Garman–Klass estimator does not take overnight dynamics into consideration, we analyze daily logarithmic returns defined as $r_t = \log(C_t) - \log(O_t)$ as well as the standardized returns $r_{st,t} = \frac{r_t}{\widehat{\sigma_{GK,t}}}$. We find that the raw returns are fat-tailed and positively skewed while the standardized returns are very close to having normal tails. This result is supported by Jarque-Bera test – the raw returns are not normally distributed but the standardized returns are very close to being normally distributed. Based on Ljung-Box test, we find no significant autocorrelations in the first thirty lags for the standardized returns. Further, we analyzed the distributional properties of volatility process and with support of the QQ plots and Jarque-Bera tests, which are not shown here, we uncover that the process of volatility is far from being normally distributed whereas its logarithmic transformation and the increments of logarithmic volatility are approximately normally distributed. Based on these findings, we propose the first three Claims[1]:

Claim 1 *Logarithmic open–close returns are uncorrelated and normally distributed with time varying volatility* $\approx N(0, \sigma_t)$.

Claim 2 *Logarithmic volatility is close to being normally distributed with mean* $\mu_{\log \sigma}$.

Claim 3 *Increments of logarithmic volatility are close to being normally distributed* $\approx N(0, \sigma_\Delta)$.

Now, we focus on an essential question of stationarity of the series. The results for ADF and KPSS tests for various lags are summarized in Table 1. We use lags 1, 10 and 100 to control for practically

[1] All "Claims" presented in this paper should be taken as approximate results. Nevertheless, we show later in the text that these "Claims" can be used to construct the series which strongly resemble the basic stylized facts of the financial returns.

	KPSS(1)	KPSS(10)	KPSS(100)	ADF(1)	ADF(10)	ADF(100)
$AAPL$	43.8283^{***}	11.2969^{***}	1.7515^{***}	-18.8717^{***}	-6.3622^{***}	-2.2830
IBM	20.5034^{***}	4.7213^{***}	0.7410^{***}	-15.0817^{***}	-5.7168^{***}	-3.1221^{**}
$MSFT$	21.2672^{***}	5.0439^{***}	0.7780^{***}	-16.1605^{***}	-5.9202^{***}	-2.6700^{*}
$\log AAPL$	57.5655^{***}	13.8110^{***}	1.9881^{***}	-16.5976^{***}	-5.6643^{***}	-1.7975
$\log IBM$	26.2098^{***}	5.9606^{***}	0.8626^{***}	-14.8127^{***}	-5.4914^{***}	-2.6609^{*}
$\log MSFT$	24.8671^{***}	5.7543^{***}	0.8176^{***}	-15.9073^{***}	-5.2080^{***}	-2.2931
$\Delta \log AAPL$	0.0012	0.0055	0.0391	-62.8961^{***}	-24.8072^{***}	-7.7096^{***}
$\Delta \log IBM$	0.0009	0.0038	0.0241	-63.5708^{***}	-23.2364^{***}	-6.7480^{***}
$\Delta \log MSFTL$	0.0015	0.0072	0.0431	-64.9717^{***}	-24.9563^{***}	-7.1949^{***}

Table 1: *Stationarity tests for* $\widehat{\sigma_{GK}}$. KPSS with the null hypothesis of stationarity and ADF with the null hypothesis of a unit root. *, ** and *** for significance at 10%, 5% and 1% significance level, respectively.

no memory, short memory and long-term memory, respectively. The results are quite straightforward. Firstly, volatility process is neither stationary not an evident unit root process. Secondly, the same is true for the logarithmic transformation of volatility. Moreover, the results are much stronger here as the series are very close to being normally distributed which is assumed for the tests. Interestingly, after controlling for long-range dependence (100 lags), we cannot reject the unit root of the series. This indicates that after controlling for long-range dependence in the increments, we cannot reject unit root for the logarithmic volatility. Thirdly, increments of the logarithmic volatility are asymptotically stationary even after controlling for long-term memory. Based on these findings, we propose three other Claims:

Claim 4 *Volatility and logarithmic volatility are non-stationary.*

Claim 5 *Logarithmic volatility is close to a unit root process after controlling for long-term memory.*

Claim 6 *Increments of logarithmic volatility are asymptotically stationary.*

Therefore, it is needed to analyze the increments of logarithmic volatility and its correlation structure. The autocorrelation and partial autocorrelation functions, not shown here, share a common pattern for all three analyzed series – strongly negative autocorrelation at the first lag for ACF which vanishes for further lags, and negative partial autocorrelations which decay slowly to zero for PACF. This is indicative for two possible processes – a strong MA(1) process or a strongly anti-persistent ARFIMA(0,d,0) process [9, 4]. Debowski [4] actually shows that ARFIMA processes can be generalized so that we obtain stationary and invertible processes even for anti-persistent processes with $d \in (-1, 0)$, i.e. $H \in (-0.5, 0.5)$. To test for anti-persistent processes while still controlling for short-term memory as well as potential over-differencing[2], we use modified rescaled range analysis and rescaled variance analysis with moving block bootstrap p-values for the null hypothesis of no anti-persistence with AR(1) and MA(1) processes in pre-whitening and post-blackening procedures. The results for $\xi = 0, 1, 2, 5, 10, 15, 30, 50, 75, 100$ are summarized in Table 2. We observe that the results for both AR(1) and MA(1) controls are practically the same – for short to medium lags, we reject the "no anti-persistence" null hypothesis, while for long lags, we do not. However, it is hard to distinguish between short and long-term memory for such long lags so that we treat the series as anti-persistent. For further discussion of the issue, see [1]. Based on these results, we propose the last Claim:

Claim 7 *Increments of logarithmic volatility are strongly anti-persistent with* $d \in (-1, 0)$.

4 Simulations

Based on all the Claims we made, we now try to reconstruct the series with the observed statistical properties and observe whether these are in hand with the real financial series. To do so, we need to estimate three parameters – $\mu_{\log \sigma}$, σ_{Δ} and d. We then simulate the series of logarithmic returns in the following way. First, we simulate ARFIMA(0,d,0) process[3] for increments of logarithmic volatility with a standard deviation of σ_{Δ}. Second, we integrate the series and adjust it so that the average value of the integrated series is $\mu_{\log \sigma}$. Third, we take the exponential of the integrated series to get the series of

[2] Since the unit root tests have low power when too many lags are taken into consideration, it is possible that taking the first differences of potentially spuriously detected unit root process imposes a strong moving average process, MA(1)

[3] We choose ARFIMA(0,d,0) because we need an anti-persistent Gaussian process which allows for strong anti-persistence. ARFIMA is an obvious and logical choice.

the time-dependent standard deviation. In the last step, we use the standard deviation for uncorrelated normally distributed series with zero mean.

ξ	V_A	M_A	V_I	M_I	V_M	M_M	V_A	M_A	V_I	M_I	V_M	M_M
0	0.000	0.000	0.000	0.000	0.000	0.000	0.008	0.000	0.005	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.004	0.001	0.000	0.003
2	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.004	0.000	0.000	0.001
5	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.002	0.000	0.000	0.003
10	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.003	0.000	0.001	0.002
15	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.003	0.000	0.000	0.002
30	0.009	0.000	0.004	0.000	0.000	0.000	0.021	0.000	0.007	0.000	0.000	0.005
50	0.071	0.000	0.015	0.000	0.006	0.003	0.103	0.001	0.022	0.000	0.009	0.022
75	0.393	0.015	0.004	0.000	0.028	0.018	0.405	0.050	0.003	0.000	0.043	0.048
100	0.770	0.068	0.009	0.000	0.161	0.083	0.714	0.193	0.004	0.000	0.155	0.139

Table 2: *Anti-persistence tests – bootstrapped p-values for the null of "no anti-persistence" controlling for AR(1) [left side] and MA(1) [right side] process.* ξ stands for the number of lags taken into consideration for standard deviation S^M for V and M statistic. Notation "A", "I" and "M" stand for AAPL, IBM and MSFT, respectively

The average logarithmic volatility $\mu_{\log\sigma}$ is -3.9962, -4.4959 and -4.3439 for AAPL, IBM and MSFT, respectively. The standard deviation of the increments of the logarithmic volatility σ_Δ is 0.4817, 0.4388 and 0.4463 for AAPL, IBM and MSFT, respectively. The biggest issue is the estimation d because majority of d and H estimators are constructed primarily for persistent processes with $d > 0$, i.e. $H > 0.5$, and their finite sample performance for anti-persistent processes has not been seriously discussed in the literature yet. To overcome this issue, we analyzed the simulations for $-0.9 \leq d \leq -0.1$ with a step of 0.1. The other two parameters are set to $\mu_{\log\sigma} = -4.5$ and $\sigma_\Delta = 0.45$.

Due to lack of space, we only show the simulations for the process which resembles the real-world financial series the best, i.e. $d = -0.5$, in Fig. 1. In the figure, we show the simulated standardized returns, simulated standard deviation process and autocorrelation function of the absolute returns. We observe that the simulations mimic the financial stylized facts remarkably – returns are uncorrelated (not shown here), volatility clustering is apparent, heavy tails are obvious and persistence of the absolute returns is visible as well (while the level of autocorrelations is close to 0.2 for all presented lags, which is exactly what is observed for empirical series [3]).

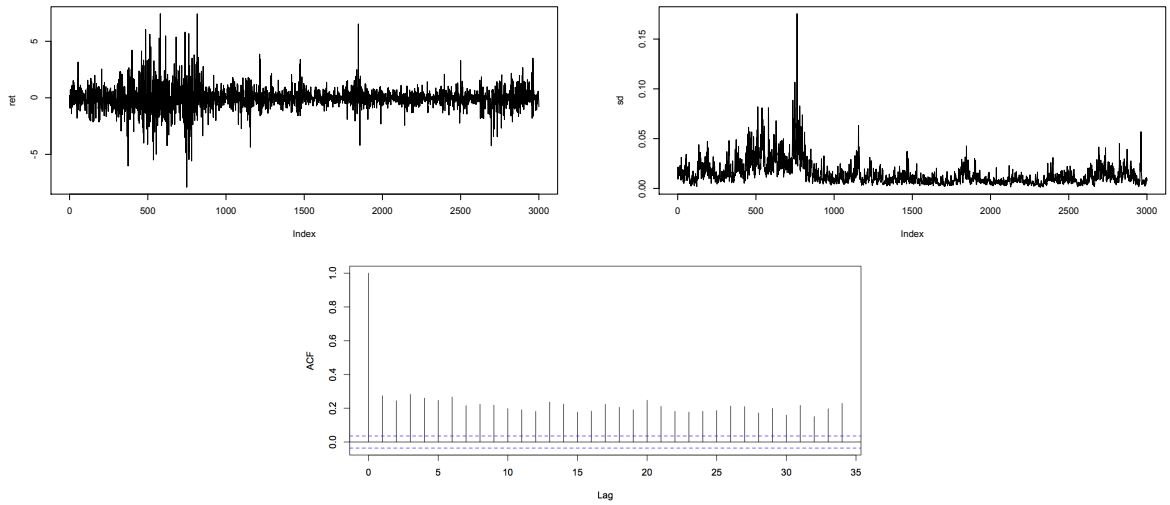


Figure 1: *Simulated process with* $\mu_{\log\sigma} = -4.5$, $\sigma_\Delta = 0.45$ *and* $d = -0.5$. *(left)* Standardized returns, *(right)* standard deviation process, and *(bottom)* autocorrelation function of absolute returns.

5 Conclusions

We introduced the alternative paradigm to volatility modeling of financial securities. On the example of three stocks of highly capitalized companies, we showed that volatility process is non-stationary and its logarithmic transformation together with logarithmic increments are approximately normally distributed. Further, the increments have been shown to be highly anti-persistent. Together with the assertion that logarithmic returns are normally distributed, and uncorrelated with time-varying volatility, we proposed the new returns-generating process. Note that the whole procedure is based on empirical observations without any limiting assumptions. We are able to construct the returns series which remarkably mimic the real-world series and posses the standard stylized facts – uncorrelated returns with heavy tails, strongly autocorrelated absolute returns and volatility clustering. Therefore, the proposed methodology opens a wholly new field in research of financial volatility. As this paper rather introduces the framework, there are many possibilities for further research in the field.

Acknowledgements

The support from GAUK 118310, GACR 402/09/0965, and SVV 265 504 are gratefully acknowledged.

References

[1] J. Beran. *Statistics for Long-Memory Processes*, volume 61 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York, 1994.

[2] R.-Y. Chou, H.-C. Chou, and N. Liu. *Handbook of Quantitative Finance and Risk Management*, chapter Range Volatility Models and Their Appliations in Finance, pages 1273–1281. Springer, 2010.

[3] R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

[4] L. Debowski. On processes with hyperbolically decaying autocorrelations. *Journal of Time Series Analysis*, 32:580–584, 2011.

[5] D. Dickey and W. Fuller. Estimators of autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74:427–431, 1979.

[6] B. Efron. Bootstrap methods: Another look at the jacknife. *Annals of Statistics*, 7:1–26, 1979.

[7] M. Garman and M. Klass. On the estimation of security price volatilities from historical data. *Journal of Business*, 53:67–78, 1980.

[8] L. Giraitis, P. Kokoszka, R. Leipus, and G. Teyssiere. Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics*, 112:265–294, 2003.

[9] J. Hosking. Fractional differencing. *Biometrika*, 68(1):165–176, 1981.

[10] H.E. Hurst. Long term storage capacity of reservoirs. *Transactions of the American Society of Engineers*, 116:770–808, 1951.

[11] D. Kwiatkowski, P. Phillips, P. Schmidt, and Y. Shin. 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*, 54:159–178, 1992.

[12] A. Lo. Long-term memory in stock market prices. *Econometrica*, 59(5):1279–1313, 1991.

[13] S.-H. Poon and C. Granger. Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, 41:478–539, 2003.

[14] S.-H. Poon and C. Granger. Practical issues in forecasting volatility. *Financial Analysts Journal*, 61(1):45–56, 2005.

[15] V. Srinivas and K. Srinivasan. Post-blackening approach for modeling dependent annual streamflows. *Journal of Hydrology*, 230:86–126, 2000.

Comparing neural networks with other predictive models in artificial stock market

Jiří Krtek [1]

Abstract. A new way of comparing models for forecasting was created. The idea was to create a simple game in which the individual compared models would compete against each other. Therefore, inspired by the heterogeneous agent models an artificial market was created. Compared models act in the artificial market as forecasting strategies of agents who trade on the market. There are traded one risky asset paying a dividend and one risk-free asset in the artificial market. The way how agents trade (buy or sell risky asset) affects the price of risky asset, which in turn influences their expectations and therefore their subsequent decisions whether to buy or sell. Moreover, each agent can recalculate parameters of his strategy, if he is not satisfied with its performance. So the forecasting strategies and the artificial market evolve side by side. It remains only to add that the winning model is the one which earns the most money.

Keywords: Neural networks, ARMA, artificial market.

JEL classification: C23, C45, C53, G17
AMS classification: 91B26

1 Introduction

In this paper are compared ARMA models as classic delegates of linear models, neural networks as delegates of non-linear models and other simple predictive models. The forecasting models are not compared using real data as it is customary to do. Accuracy of forecasts or their standard deviations are not calculated. Instead, a new, unconventional method for comparing strategies was created. The models are compared in an artificial stock market. Traders, or agents, in the market use the aforementioned models as forecasting strategies. The best model among them is simply the one that earns the most money. It is also important to create an artificial stock market that has similar characteristics to the real market. The characteristics of the artificial market are therefore also studied and the artificial market is built to be conformable to the real one.

The structure of the market is inspired by several papers (see next section), most features are derived but several are products of own invention. The aim of this work is to create new way of comparing forecasting strategies and to compare primarily performance of linear and non-linear models.

2 Artificial stock market

The model which simulates market environment was inspired mostly by [2], [7], [6] and [8]. Two assets are traded in the market and no transaction costs. The first is a risk-free asset which is perfectly elastically supplied and has a piecewise linear rate of return r_t. The dynamics of the risk-free asset is product of own invention and is given by the following formula

$$r_{t+1} = r_t + \varepsilon_{r,t}\eta_{r,t}\Delta_r, \tag{1}$$

where $r_0 = 5 \times 10^{-5}$, $\Delta_r = 5 \times 10^{-6}$ are constants, $\varepsilon_{r,t}$ is alternatively distributed random variable ($\varepsilon_{r,t} = 1$ with probability $p_\varepsilon = 10^{-2}$, otherwise $\varepsilon_{r,t} = 0$) and $\eta_{r,t}$ is another alternatively distributed random variable ($\eta_{r,t} = 1$ with probability $p_\eta = 0.5$, otherwise $\eta_{r,t} = -1$). There is one more condition

[1] The Faculty of Mathemetics and Physics by Charles Univesiry in Prague, Department of Probability and Mathematical Statistics, Sokolovská 83, Praha 8, e-mail: krtek@karlin.mff.cuni.cz

stating that r_t is always positive, i.e. it holds that $r_t > 0$, $\forall t$. It should be noted that agents don't have to forecast the change of r_t because it is always known ahead. Thus r_t can be interpreted as short interest rate. Traders make a deposit with this safe short interest rate when they have redundant money.

The second asset is a risky stock whose price at time t is denoted by p_t. The risky stock can be divided into endlessly small pieces and pays dividend d_t at time t; the dividend is chosen (as in [6]) to follow the stationary AR process

$$d_{t+1} = \bar{d} + \rho(d_t - \bar{d}) + \varepsilon_{d,t+1}, \tag{2}$$

where $\rho = 0.95$ is chosen this way to provide a persistent process which is still stationary, $\bar{d} = 0.05$ is a constant and $\varepsilon_{d,t} \sim N(0, 10^{-2})$.

There are 8 groups of traders. Most of the traders predicts the future changes of price and future dividend itself. So they don't predict the future price of the risky asset itself, but a change from the last known value because it is assumed (in accordance with the real market) that the price process is not stationary. All traders forecast the future price (or price change) and dividend using their lagged values. As in [2] and many other heterogeneous-agent models, it is assumed that all traders maximize their expected utility function. It is also assumed that all traders have the same constant absolute risk aversion utility function U, with the same risk aversion parameter $\lambda = 1$. Let $W_{i,t}$ denote the wealth of trader i at time t. Furthermore, let $h_{i,t}$ be the number of agent i's shares at time t, then trader's goal is to maximize the expected utility at time $t+1$ given information up to time t over his number of shares, i.e.,

$$\max_{h_i,t} E[U(W_{i,t+1})|I_t] = E_{i,t}[U(W_{i,t+1})], \tag{3}$$

where I_t denotes the information set available at time t. Under the assumption of exponential CARA utility function and the Gaussian distribution of forecasts, the optimal number of shares at time $t+1$ is given by the following ratio

$$h^*_{i,t+1} = \frac{E_{i,t}[p_{t+1} + d_{t+1}] - (1+r)p_t}{\lambda \sigma^2_{i,t}}, \tag{4}$$

where $\sigma^2_{i,t}$ is the conditional variance of $p_{t+1} + d_{t+1}$ given I_t. Traders differ only in their forecasting strategies, i.e., the way they calculate $E_{i,t}[p_{t+1} + d_{t+1}]$. The conditional variance of $p_{t+1} + d_{t+1}$, i.e., the term $\sigma^2_{i,t}$ is estimated in the same way by all traders. Following [2], the estimate of conditional variance is given by

$$\sigma^2_{i,t} = (1-\theta)\sigma^2_{t-1|n} + \theta(p_t + d_t - E_{i,t-1}[p_t + d_t])^2, \tag{5}$$

where $\theta = 0.01333$, $n = 10$

$$\sigma^2_{t|n} = \frac{\sum_{j=0}^{n-1}[P_{t-j} - \bar{P}_{t|n}]^2}{n-1} \tag{6}$$

with

$$\bar{P}_{t|n} = \frac{\sum_{j=0}^{n-1} P_{t-j}}{n}. \tag{7}$$

2.1 Price evolution

The model of price evolution is inspired by [2] and [8]. According to the previous section, $h^*_{i,t}$ will be used for the desired number of risky shares at time t, while $h_{i,t}$ will denote the actual number of shares held. Let $b_{i,t}$, be the number of shares that trader i would like to buy and let $a_{i,t}$ be the number of shares which he would like to sell, i.e.,

$$b_{i,t} = \begin{cases} h^*_{i,t} - h_{i,t-1}, & h^*_{i,t} \geq h_{i,t-1}, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

$$a_{i,t} = \begin{cases} h_{i,t-1} - h^*_{i,t}, & h^*_{i,t} < h_{i,t-1}, \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

Let N be the total number of traders. If $B_t = \sum_{i=1}^{N} b_{i,t}$ denotes the overall demand for the risky asset and $A_t = \sum_{i=1}^{N} a_{i,t}$ the overall supply of the risky asset, then the number of shares held by trader i

at time t is dependent on the number of shares he/she wanted to buy/sell. Traders' offers and demands are first sorted in descending order, subsequently the offers, respectively demands are satisfied as long as there are some demands, respectively offers. So the bigger volume of an offer/demand the bigger probability that it will be satisfied.

The price adjustment is based on the excess demand $B_t - A_t$, it is inspired by (but not completely adopted from) [2] and is given by

$$p_{t+1} = p_t \cdot [1 + \beta(B_t - A_t) \cdot \min(A_t, B_t) \,/\, Total_{shares}] + \eta_{p,t+1}\varepsilon_{p,t+1}, \tag{10}$$

where $\eta_{p,t}$ is alternatively distributed random variable, $\eta_{p,t} = 1$ with probability $p_p = 2 \times 10^{-2}$ and $\eta_{p,t} = 0$ otherwise, and $\varepsilon_{p,t}$ is the noise (i.i.d. random variables from $N(0, 4)$) added in order to represent other traders potentially present in the market, and β is some function of excess demand. Following [2], β was set to

$$\beta(B_t - A_t) = \begin{cases} \tanh(\beta_1(B_t - A_t)), & B_t \geq A_t, \\ \tanh(\beta_2(B_t - A_t)), & B_t < A_t. \end{cases} \tag{11}$$

The parameters β_1 and β_2 were set to 10^{-4} and 2×10^{-4} respectively. It is clear that the price does not change – except the possible noise – when there is no demand for risky asset or no supply of it which is in accordance with the real market.

2.2 Forecasting strategies

Traders want to forecast future change of the price and future dividends. As was previously stated, this work compares neural networks with more traditional models used for prediction. Traders therefore use neural network, ARMA or another model as their forecasting strategy. This is the only thing that differentiates one trader from another. Each trader uses the same forecasting strategy for both future price (or log-return, i.e. change of logarithm of the price) prediction and future dividend prediction.

The first group of agents uses AR models as their forecasting strategy for a prediction of the log-return and the dividend. Parameters of the AR models are estimated by conditional maximum likelihood (see [4] for details). Traders of the second type forecast the same future values with the aid of ARMA models estimated by conditional maximum likelihood again. Traders of the third and fourth type uses a feed-forward neural network with 3 and 4 layers respectively to forecast the same quantities. The Levenberg–Marquardt algorithm and backpropagation are used for searching for optimal parameters, see Section 2.3 for details. The fifth group of traders utilizes the Elman recurrent networks for forecasting the log-returns and dividends again. The Levenberg–Marquardt algorithm is used for searching for optimal parameters. For all NN models, the activation function of hidden neurons is tangent hyperbolic and the activation function of output neuron is linearity. For more information about feed-forward neural networks and Elman's networks see [5], [1], [9] or [3].

There are three more types of forecasting strategy. The first of them is strategy which employs an average of last realized values to predict the future price itself and future dividend. The second is so-called Naive forecasting strategy which can be viewed as special case of the previous. It simply predicts the future value by the last known. This forecasting strategy can be viewed as the basic for comparison with other more advanced strategies. Traders with the last strategy make random predictions of future price changes and future dividends.

Although some tables presented in Subsection 3.2 contain also results for traders utilizing the average and random forecasting strategy, these traders fulfill special roles in the artificial market. The purpose of random traders is to bring liquidity in the market. Traders employing average serve as an antipole to all the trend-chasers in the market, i.e. traders forecasting with AR or ARMA models or any kind of neural network, so they essentially fulfill the role of fundamentalists. They should force the price to return back. As these two types of traders should somehow stabilize the market (average) or should support trading (random) the number of traders with these two strategies is bigger than the number of other traders. Specifically, there is 280 random traders and 40 traders using averages, while each other group of traders has 4 only traders. Some tables in Subsection 3.2 don't contain results for the strategies fulfilling special roles because the main goal of this work is to compare NNs and ARMA models. Moreover in some cases it don't make sense to present results for these strategies.

2.3 Learning of traders

The learning is partially own invention and partially taken from [2]. A part of trader learning from [2] is adopted and part is left out. Every trader knows his wealth as well as the wealth of all other agents in this artificial market. After the given period of time $k = 400$, each trader counts the difference between his/her present wealth and wealth at time $t - k$, i.e., $W_{i,t} - W_{i,t-k}$. This indicates how much money the trader has earned or lost. Everyone consequently gets a rank $R_{i,t}$ according his difference of wealth. The quantity

$$r_{i,t} = \frac{R_{i,t}}{N} \tag{12}$$

is then the probability that agent i recounts parameters of his forecasting strategy due to *pressure of society*. The smaller his rank the smaller the probability that he learns new parameters of his strategy (because he is satisfied with the current parameters). Each trader can also recounts his strategy on the basis of growth rate of his wealth over the previous period. Therefore, even if trader i does not recount his strategy because of *pressure of society*, there is still chance that he will do it for a different reason. Specifically, because he is not satisfied with himself. Let

$$\chi^k_{i,t} = \frac{W_{i,t} - W_{i,t-k}}{|W_{i,t-k}|} \tag{13}$$

be trader i's growth rate of wealth over the previous period. It in fact measures how effective trader i's strategy was. Then

$$s_{i,t} = \frac{1}{1 + \exp\{\chi^k_{i,t}\}} \tag{14}$$

is the probability that trader i will recount his strategy because of its low efficiency. If trader i recounts parameters of his strategy at the end of the tth day, he will learn on the data of length $m = 300$ (common to all traders).

Except aforementioned learning, the traders using the forecasting strategies based on feed-forward neural networks can recount the parameters of their strategies each time after the new log-return and the new dividend amount is known. In that time, they already know the correct log-return and dividend they forecasted before so they can update the parameters of their forecasting strategies simply by backpropagation with momentum.

3 Simulation

Fourteen simulations with the same setups were performed. Traders always learn at the beginning of the simulation on simulated data of length 150. The dividend process was generated according to (2). For price, its mean was set to 1000 and the log-returns were taken from GARCH(1, 1) process with Gaussian innovations. Traders estimated parameters of their strategies on these simulated data. After this initial learning the traders were given initial amount of money (800) and shares (0.15). The last values of simulated dividend and price served as initial values to the market simulation. Traders were let to trade for 200 time steps. After this period all agents recalculate (learn) the parameters of their forecasting strategies, the realized price series (200 time steps) were discarded and the same initial amount of money and shares were given to all of them. The same procedure was applied again, i.e. 200 time steps of trading, recalculation of parameters, price series discarded and initial money and shares allocation. Then the competition (described in this paper) having aforementioned 1170 time steps started. The previous two procedures were performed in order that agents learn the parameters of their strategies on data arisen from their trading.

3.1 Verifying stylized facts

To get more realistic results, the constructed market should have similar characteristics as the real markets have. The characteristics of real markets (also called stylized facts) are for example heavy tail distribution of returns and their volatility clustering or that prices follow random walk. As can be seen from the numbers presented below, some stylized facts hold in the artificial market and some not. The results of testing are presented in the Table 1. The first tested stylized fact (column *adf*) was that the price

series have a unit root. This hypothesis was tested by augmented Dickey–Fuller test. The Kwiatkowski–Phillips–Schmidt–Shin test (column *kpss*) was used to test the second stylized fact – the null about level stationarity of log-returns. The hypothesis that the log-returns are uncorrelated in time – third stylized fact (column *B–P*) – was tested by Box–Pierce test. The last stylized fact tested was that log-returns have distribution with fat tails. It was tested whether the log-returns have Gaussian distribution and also their kurtosis was calculated. The goal of this analysis was to show that the hypothesis about Gaussian distribution of log-returns in the artificial market is rejected. The test used for this purpose was Jarque–Bera test, the column *J–B*. The kurtosis indicates fat tails when it is higher than 3 (the kurtosis of Gaussian distribution) which is the reason why it was calculated, column *kurt*. The last row of the Table 1, named *summary*, contains proportion of null rejections on 5% confidence level for first four columns. For the last column (*kurt*) it contains the proportion of excess kurtosis, i.e. the case when the calculated kurtosis is higher then 3.

simulation no.	adf	kpss	B–P	J–B	kurt
1	0.65	0.10	~ 0	~ 0	5.0
2	0.95	0.06	~ 0	~ 0	4.7
3	0.87	0.04	~ 0	~ 0	3.9
4	0.69	0.09	~ 0	~ 0	4.6
5	0.58	0.10	~ 0	~ 0	4.2
6	0.77	0.10	~ 0	~ 0	3.8
7	0.93	0.09	~ 0	~ 0	4.4
8	0.48	0.03	~ 0	~ 0	3.6
9	0.39	0.01	~ 0	~ 0	4.9
10	0.92	0.10	~ 0	~ 0	4.4
11	0.70	0.10	~ 0	~ 0	5.6
12	0.53	0.10	~ 0	~ 0	3.6
13	0.95	0.01	~ 0	~ 0	3.9
14	0.80	0.10	~ 0	~ 0	3.7
summary	0 %	28.6 %	100 %	100 %	100 %

Table 1: Results of stylized facts testing.

3.2 Comparing strategies

When comparing agents, the criterion of their success was the amount of money they earned. Specifically, it was the ratio of last and initial wealth of the particular traders. The results of the 14 simulations are presented in the Table 2 below. The very interesting phenomenon is that the traders utilizing Elman networks won every simulation with only one exception (simulation no. 11) where the traders with ARMA models as their forecasting strategy won. The best trader in all simulations was utilizing Elman networks. So the 11th simulation was won by traders using ARMA models in average, however they all were beaten by one of the traders forecasting with Elman networks. Except the presented results other more common statistics were also studied, e.g. root mean square error and sign test of price predictions. Traders with Elman networks as forecasting strategy had the lowest RMSE in most of simulations. Regarding the sign test, the Elman netowrks were the best strategy in the direction estimation in ten out of the 14 cases. The results are also interesting in the following feature. Although the FF1L and FF2L strategies had much better results with respect to the sign test than AR models, the results of all these three strategies are pretty the same for the ratio of final and initial wealth (see Table 2). This might seem quite strange, but it is exactly the reason why the competition, i.e. the market simulation, was performed.

4 Conclusions

It was created an artificial market in order to compare forecasting strategies in it. The purpose of it was to show a new way of comparing strategies. Except one stylized fact, the requirement of unpredictable

sim. no.	AR	ARMA	Avg	Elman	FF1L	FF2L	Naive	Rnd
1	1.035	1.107	0.973	1.110	1.061	1.043	1.077	1.068
2	1.030	1.117	0.965	1.161	1.052	1.067	1.081	1.066
3	1.051	1.146	0.983	1.178	1.024	1.031	1.044	1.058
4	1.016	1.088	0.943	1.104	1.032	1.051	1.052	1.042
5	1.057	1.183	0.977	1.189	1.030	1.048	1.044	1.059
6	1.053	1.110	0.974	1.202	1.056	1.035	1.038	1.055
7	1.031	1.143	0.955	1.177	1.017	1.032	1.055	1.050
8	1.037	1.120	0.958	1.164	1.045	1.049	1.049	1.048
9	1.032	1.093	0.960	1.187	1.044	1.047	1.056	1.057
10	1.060	1.121	0.993	1.191	1.091	1.081	1.087	1.084
11	1.038	1.142	0.970	1.102	1.039	1.044	1.074	1.066
12	1.044	1.130	0.970	1.186	1.036	1.020	1.038	1.050
13	1.040	1.103	0.954	1.186	1.045	1.039	1.041	1.043
14	1.043	1.105	0.973	1.231	1.021	1.034	1.035	1.047
average	1.040	1.122	0.968	1.169	1.042	1.044	1.055	1.057

Table 2: The average ratio of final and initial wealth for particular groups of traders. AR = AR models, ARMA = ARMA models, Avg = moving average, Elman = Elman networks, FF1L = feed-forward NN with 1 hidden layer, FF2L = feed-forward NN with 2 hidden layers, Naive = naive strategy, Rnd = random forecast

(or at least uncorrelated) log-returns, the artificial market seems to have similar characteristics as the real one. The goal of this paper was to compare the NNs and ARMA models. The Elman networks completely dominate in the presented contest "the winner is the one who earns the most money" as well as in the customary statistics (linked to the outputs of the market simulation) such as RMSE and sign test.

Acknowledgements

I would like to thank my supervisor Miloslav Vošvrda for helpful comments. The work was supported by grant of GA ČR P402/12/G097. The Bernard Bolzano Endowment Fund is also greatly acknowledged.

References

[1] Boden, M.: *A Guide to Recurrent Neural Networks and Backpropagation*. In the Dallas project, 2002.

[2] Chen, S., and Yeh, C.: *Genetic Programming in Agent-Based Artificial Markets: Simulations and Analysis*.

[3] Dorffner, G.: Neural Networks for Time Series Processing, *Neural Network World* **6** (1996), 447–468.

[4] Hamilton, J. D: *Time Series Analysis*. Princeton University Press, 1994.

[5] Haykin, S.: *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1999.

[6] LeBaron, B.: *Building the Santa Fe Artificial Stock Market*. Brandeis University, 2002.

[7] Matassini, L., and Franci, F.: On Financial Markets Trading, *Physica A: Statistical Mechanics and its Applications* **3–4** (2001), 526–542.

[8] Palmer, R. G., and Arthur, W. B., and Holland, J. H., and LeBaron, B., and Taylor, P.: Artificial Economic Life: A Simple Model of a Stockmarket, *Physica D* **75** (1994), 264–274.

[9] Šíma, J., and Neruda, R.: *Teoretické Otázky Neuronových Sítí*. Matfyzpress, Praha, 1996.

Multiple-criteria assessment of edges in vehicle routing problems

Petr Kučera[1]

Abstract. Let us consider, in spite of the paper title, an ordinary vehicle routing problem with one optimization criterion. Some methods, however, use slightly different edge evaluations, based of course on the original criterion. For instance, the famous Clark and Wright method uses the savings. If we have two (or more) such evaluations, there rises a question how to apply them in one time. Should we regard as good evaluated edges with good assessments by both evaluations or is it sufficient to be good according to one evaluation only? Or should we take rating by average?
The answer may help in right setting of genetic algorithms. The chromosomes in genetic algorithms are also edge evaluations. If we know the answer to the question put above, we can e.g. decide whether it is more suitable to define a vocation to be a good edge as dominant or recessive, or select individuals for next solution breeding in a better way.
In this contribution, we test and compare the results of the multiple-criteria approaches described above on several test cases.

Keywords: vehicle routing problem, combinatorial optimization, multiple-criteria evaluation, genetic algorithm.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

In practice we often come across a task how to distribute certain material from/to the central city (point, place) to/ from a finite number of cities (places) using a circular tour. When the capacity of one vehicle is sufficient, we endeavour to find a tour which passes exactly once through each of the given cities and the total length of the tour is minimized. This task is called Traveling Salesman Problem (TSP). In the opposite case, when a vehicle has not enough capacity, we have to design more than one circular tour, i.e. to use more than one vehicle or the vehicle must make more tours. Such tasks where more than one vehicle is necessary for the transportation due to various reasons are called Vehicle Routing Problems (VRP). We will observe a task with vehicle capacity limitation as described above and, besides this, with all the vehicles having the same capacity (i.e. the vehicle fleet is homogenous). There is one central city and each vehicle both starts and finishes its route in the central city (i.e. the VRP is closed). Such VPRs are also sometime called Multiple-tour Travelling Salesman Problems (cf. e.g. [4] or [5]).

Let us introduce some notation. The central city will receive index 0 and the other cities numbers from 1 to n. The cost matrix will be denoted by $\mathbf{C}$ (and so single costs c_{ij}, $i, j = 0, \ldots, n$).

Both TSP and VRP belong among the NP-hard problems, for which there is no efficient algorithm find their theoretical optimum. So the only way how to obtain some solution efficiently or, in a reasonably short time, is to use some heuristics (approximation methods) which give only "good" or "close to optimal" solution, not the exact optimum. A large number of such heuristics has been designed. Let us mention at least several of them. One of the first heuristics is the savings method by Clarke and Wright [2]. Also Habr [3], Czech scientist, regarded as a founder of the Czech systems school, is the author of several other early heuristics based on so called frequencies. On the other hand, one of recent approaches to solving the NP-hard problems appears in genetic algorithms (e.g. [1] for TSP, [7] and [8] for VRP, [6] for other tasks).

These examples have not been chosen by chance. Although the VRP is a task with one optimization criterion given by the cost matrix where the costs evaluate straight routes (edges) between single pairs of cities, all the methods mentioned above create its own edge assessments. The aim of this paper is to demonstrate on several test cases, if we have two (or more) such edge evaluations, how to combine them to obtain the possible best solution.

[1] Czech University of Life Sciences in Prague, Department of Systems Engineering, Kamýcká 129, Praha 6, kucera@pef.czu.cz.

2 General definition of an edge evaluation and tour construction based on edge evaluation

We will define evaluations (assessments) only for edges which do *not* contain (are *not* incident to) the central city. Such an evaluation forms a matrix **A** (and evaluations for single edges are a_{ij}, $i, j = 1, \ldots, n$). The solution of a given VRP will be obtained from this evaluation by the following way:

Tour construction according to the evaluation matrix

- Process the edges according to the descending order of the evaluations a_{ij} (from the best to worst) using the following rules:
- Upon adding the edge, if after adding this edge all the edges so far added to the route form the set of vertex disjoint paths and for each path the sum of the capacities of the cities lying on it does not exceed the capacity of the vehicle, then add it to the solution.
- Repeat the procedure until each city lies on some of the paths and joining arbitrary two paths the vehicle capacity is exceeded.
- In the end add the city 0 to the ends of all the paths to form cyclic routes.

In the following, we will apply this procedure using different edge assessments.

3 Savings method

The savings method is based on comparing lengths of a straight route between any two cities and a route via another fixed selected city, using the following algorithm. In case of the VRP it is natural to use the central city as the fixed one. Then the savings for all the edges are computed using the following formula:

$$s_{ij} = c_{i0} + c_{0j} - c_{ij} \tag{1}$$

Thus, the savings are an illustrative example of the edge evaluation described in the previous chapter. In [5], exactly the same procedure from the previous chapter is applied, too.

4 Habr frequencies

Another case of edge evaluation is represented by frequencies introduced by Habr and based on the comparison of an edge with all other edges (not only with those containing one selected city as in the case of the savings). The frequencies can be expressed using the following formula:

$$F_{ij} = \sum_{k=1}^{n} \sum_{l=1}^{n} (c_{ij} + c_{kl} - c_{il} - c_{kj}) \tag{2}$$

In the application for the VRP the edges containing the central city (as they occur in a randomly chosen solution with a higher probability than the others) are taken in the frequencies (2) with a bigger weight in the sum than the others. The more detailed description can be found in [5].

Let us remark that, in difference from the savings, the Habr frequencies are the minimization criterion (the lower the frequency is, the better the edge is). When handling all the edge evaluations, we will apply the normalization (see chapter 6 below, formula (3)) which transforms all of them to maximization criteria.

5 Genetic algorithms

Genetic algorithms try to mimic evolution biology principles to find solutions of complex optimization problems. There are a large number of different types of genetic algorithms, even for a particular task. The genetic algorithm operates with a population of solutions. At the start of running, this population is usually formed randomly. Each individual of the population is determined by its chromosome. During the run of a genetic algorithm, pairs of individuals (solutions) are selected to be crossed and thus new solutions arise and the population ramifies. A particular genetic algorithm comprises the rules for new individual formation when crossed, and, besides that, also the rules for selecting pairs of individuals for breeding or a "natural" selection, so that the newly born solutions are as good as possible (all the rules are usually partially stochastic). Another important feature and a strong tool of genetic algorithms is the possibility of random mutations (changing one item of the chromosome) during the crossing (although we will not observe it in this paper). Namely, it leads to the creation of new possibly good solutions significantly different from those obtained during the previous algorithm run.

For the VRP, the chromosome is actually also an evaluation of edges. In this paper, we will define and use the chromosome as an evaluation of only those edges not containing the central city (as in Chapter 2). In particular, the savings matrix, frequencies matrix, and cost matrix without the costs of the edges not containing the central city can be considered special cases of chromosomes. The VRP solutions will be derived from chromosomes by the same manner as described in Chapter 2.

The crossing of solutions in genetic algorithms is actually a case of finding a new solution based on a two-criterion evaluation of edges in the VRP. Therefore it makes sense to put a question how to carry it out in the best possible way and thus obtain high-quality chromosomes for new individual solutions.

6 Testing of multiple-criteria assessment

For testing we used exactly the same five test cases as in [5]. They were randomly generated as follows: We supposed a circle with 100 km diameter with 20 km diameter circle in the centre with the central city only. We generated 20 cities outside the smaller circle with capacities from 250 to 550 units (the capacities have equal distribution and 50 units minimal difference). Up to 4 closest cities we joint into "region" with a centre in the middle or in the biggest city. Generation was terminated when 12 regions were created. We also used vehicles with the same capacity as in [4] and [5], i.e. 2100 units.

We used the cost matrix (without the costs of the edges not incident to the central city), savings matrix, and frequencies matrix. We perform the normalization of all these three matrices using the formula

$$r_{ij} = \frac{a_{ij} - D}{H - D} \tag{3}$$

where **R** and **A** are the normalized and original matrices, respectively, and H and D are the best and worst evaluation in the original matrix **A**. This normalization converts both maximization and minimization criterion to the form where all the edges obtain evaluations between 0 and 1, the best edge gets 1 and the worst one 0.

Then, for testing the multiple-criteria (or more precisely, two-criterion) assessments, we considered all pairs of these three normalized matrices. For each of these pairs we created the following three matrices:

- the matrix, elements of which were maxima of corresponding elements of the original matrices (in terms of genetics, it simulated the case when the vocation to be a good edge is dominant, i.e. if one parent had this vocation, the descendant had it, too),
- the matrix, elements of which were minima of corresponding elements of the original matrices (it simulated the case when the vocation to be a good edge is recessive, i.e. a necessary condition for having the vocation for a descendant was that both the parents had it),
- the matrix, elements of which were average values of corresponding elements of the original matrices (both the parents contributed to their child's vocation in the same extent).

Thus, for each test case, we obtained altogether nine new assessment matrices (chromosomes of newly arising "hybrid" solutions).

Criterion	Case 1	Case 2	Case 3	Case 4	Case 5	Average
S	100.0%	100.0%	100.0%	104.6%	100.0%	-
F	100.0%	101.9%	100.0%	100.0%	101.6%	-
CxS max.	105.3%	106.6%	109.4%	106.2%	103.8%	106.3%
CxF max.	105.3%	106.6%	109.4%	113.7%	103.8%	107.8%
SxF max.	100.0%	100.7%	98.8%	104.6%	101.6%	101.1%
CxS min.	100.0%	100.0%	100.3%	104.6%	100.0%	101.0%
CxF min.	100.0%	101.9%	102.4%	110.5%	101.4%	103.2%
SxF min.	100.0%	101.2%	102.0%	104.6%	100.0%	101.6%
CxS aver.	109.1%	101.9%	103.4%	104.6%	101.4%	104.1%
CxF aver.	100.0%	100.7%	102.4%	104.7%	101.4%	101.8%
SxF aver.	100.0%	100.0%	101.6%	104.6%	101.6%	101.6%

Table 1 Test cases results

In Table 1 are results in the percentage form. We express as 100% the better one of the solutions obtained using the savings method and the frequencies approach. We use this expression in order to show how good new

solutions in comparison with the original ones are. In the first two rows there are the solutions by the savings method (S) and the frequencies approach (F) (cf. [5]), in the remaining nine rows the solutions according to the newly created assessment matrices (chromosomes). In the right-hand side column there are average results of the respective type of assessment.

First of all, we can see from the table that the combination of the costs with another criterion using the maximum values gave the worst solutions. On the other hand, the only case of a better solution than the original ones was achieved also using the maximum values by combining the savings and frequencies. When using both the minimum and average values, the results were relatively stable; very bad solutions occurred only rarely and, on the contrary, we often obtained the original solutions. However, it is interesting that in Case 4, i.e. the only case when (considering the original solutions) the frequencies gave a better solution than savings, we did not even once get (using combinations) this best solution in difference from all other cases.

7 Conclusion

Based on the results above, we can recommend for genetic algorithms designing to use maximum edge evaluation when crossing, i.e. to define the vocation to be a good edge as dominant (this will lead to the rise of new individuals of both very high and very low quality); and at the same time to select strictly high-class individuals for crossing to avoid the rise of an unnecessarily large number of poor individuals.

In contrast, for simple "manual" computation we propose to use minimum or average values as they may provide a wide spectrum of good solutions for choice for a user.

References

[1] Braun, H.: On Solving Traveling Salesman Problems by Genetic Algorithms. *Lecture Notes in Computer Science* **496** (1991), 129-133.

[2] Clarke, G., Wright, J.W.: Scheduling of Vehicles from a Central Depot to a Number of Delivery Points. *Operations Research* **12** (1964), 568-581.

[3] Habr, J.: *Jednoduché optimalizační metody pro ekonomickou praxi*. SNTL Prague, 1964.

[4] Kučera, P., Dömeová, L.: Solution of the Multipletours Traveling Salesman Problem with Vehicle Capacity Preference. . In: *Proceedings of the 21st International Conference on Mathematical Methods in Economics 2003*. Czech Society for Operations Research, Prague, 2003, 171-176.

[5] Kučera, P., Houška, M., Beránková, M.H.: Methods for the Multipletours Traveling Salesman Problem Making the Final Solution in One Time. In: *Proceedings of the 24th International Conference on Mathematical Methods in Economics 2006*. Czech Society for Operations Research, Prague, 2006, 313-316.

[6] Perzina, R.: Self-adaptive Genetic Algorithm for Solving Travelling Purchaser Problem. In: *Proceedings of the 29th International Conference on Mathematical Methods in Economics 2011*. Czech Society for Operations Research, Prague, 2011, 545-550.

[7] Potvin, J.Y., Dube, D., Robillard C.: A hybrid approach to vehicle routing using neural networks and genetic algorithms. *Applied Intelligence* **6** (1996), 241-252.

[8] Potvin, J.Y., Duhamel, C., Guertin F.: A genetic algorithm for vehicle routing with backhauling. *Applied Intelligence* **6** (1996), 345-355.

The role of foreign trade in the process of financial integration: The case of European Union countries

Zuzana Kučerová[1]

Abstract. The importance of monetary integration is increasing and the adoption of a common currency has to be accompanied by financial market integration. Using a data sample of EU member countries over the period 1994-2008 we assess the impact of the individual factors which dominated the process of financial market integration in the selected region. The empirical model is estimated by using the econometric method of panel regression. Because of the existence of reverse causality, i.e. the existence of simultaneity between the dependent variable and one of the explanatory variables (foreign trade) of the "basic" model, another equation was added into the model and this simultaneous equations model was then estimated by using the instrumental variables method and the two stage least squares method (the "new" model). We conclude that foreign trade has a positive impact on the financial integration and that increasing financial integration leads to more intensive foreign trade among these countries. However, world trade (and thus financial integration) can be undermined by limiting trade finance as a result of tightening financial regulation and supervision after a decade of loose regulation and supervision.

Keywords: financial integration, foreign trade, international investment position.

JEL Classification: C23, C36, E44, F36, F42
AMS Classification: 62P20, 62M10, 91B84

1 Introduction

The integration of financial markets contributes to the overall integration and economic growth by removing the exchange rate risk and the barriers and frictions in cross-border capital movement. [2] This allows the capital to be allocated more efficiently (Baele *et al*. [1]). Financial integration is an important factor in increasing the efficiency of a financial system and lowering the costs for business as well as for consumers. However, some negative effects can be linked with this process. Some capital may hinder the economy without barriers for capital movement, especially the so-called "hot money", which can be transferred from one country to another very quickly and without restrictions and major expenses. The process of financial integration of the past decade was associated with an unprecedented 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 [4]).

The aim of the paper is to analyse the degree of financial integration of the EU countries with the rest of the world by using quantity-based measures of financial integration derived from the countries' international investment positions with a view to the foreign trade. We investigate the relations between the country's foreign assets and liabilities (a dependent variable) and selected macroeconomic variables (explanatory variables). The paper is structured as follows. The first chapter is introduction. In chapter two, the measures of financial integration and some remarks on the previous empirical research concerning the analyses of international investment positions are presented. In chapter three, the data, periods and countries and methods used in this paper are described. In chapter four, we present and estimate the "basic" and "new" empirical model in order to analyse the degree of financial integration among the EU27, EU10, and EU17 countries and the rest of the world by using the panel regression method. Chapter five is a discussion concerning the role of trade finance in the process of financial integration. The last chapter summarises the results and brings conclusions.

[1] Faculty of Economics, VSB-Technical University of Ostrava, Dept. of Economic Policy, Sokolská tř. 33, Ostrava 1, 701 21, zuzana.kucerova@vsb.cz.
[2] See Kotlanova and Kotlan [9] for verification of other channels that growth variables, especially taxation, can influence economic growth by.

2 Financial integration

2.1 Measures of financial integration

Baele *et al*. [1] consider three categories of measures: price-based, news-based, and quantity-based measures. *Price-based measures* quantify discrepancies in prices or yields on assets, which result from the assets' geographic origins. This set of measures allows us to check the validity of the law of one price. *News-based measures* differentiate between information effects and frictions or barriers. News of a regional character is expected to have a smaller effect on asset prices than, for example, global news. *Quantity-based measures* are designated to measure the effects of frictions faced by the demand or supply on the financial markets. These measures are based on asset stocks and flows. An analysis of these data can give us information on, for example, the ease of market access or cross-border activities. This paper uses this type of measures.

2.2 Previous research

A shortage of data was the main obstacle to analyses of international investment positions several years ago. Thanks to the International Monetary Fund and its statistic databases we are now able perform these analyses. The paper focuses on changes in a country's international investment position, especially in foreign assets and liabilities.[3] Our paper follows the line of this research, studying a set of factors affecting foreign assets and liabilities. Lane and Milesi-Ferretti [11] created a methodology to produce a unique data set containing an estimation of foreign assets and liabilities for a large set of industrial and developing countries for the last three decades. This data set has enabled to analyse net foreign assets in a more complex way. It is one of the first attempts to study the foreign assets and liabilities. Lane and Milesi-Ferretti [12] focused on variables influencing net foreign assets. They study the effects of changes in output, public debt, and demographic factors on net foreign assets. Lane and Milesi-Ferretti [13] examine the cross-country and time-series variation in the size of international balance sheets. They study the relation between foreign assets and liabilities and a set of various regressors (GDP per capita, trade openness, external liberalisation, financial depth, stock market capitalisation, privatisation revenues etc.). Lane and Milesi-Ferretti [14] explore net foreign assets in Central and Eastern European countries for the last ten years, taking into account the composition of their international investment positions. They concentrate especially on foreign direct investments, portfolio investments, and external debt financing. Goldberg [6] explores patterns in the international exposures of U.S. banks, regressing changes in foreign assets and liabilities and changes in domestic and foreign real interest rates and real GDP in the United States. Buch *et al*. [3] analyse whether foreign bank assets react to macroeconomic shocks as the economic theory predicts and how valuation effects influence changes in foreign banks assets. Kose *et al*. [8] 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 a country's financial openness (the sum of financial assets and liabilities relative to the nominal GDP).

3 Data and methods

3.1 Data

The data used to calculate the measures of financial integration are from the International Monetary Fund (IMF) International Financial Statistics (IFS) online database, specifically a category called the international investment position (IIP). 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 [15]). This database contains data for the period 1970-2007 for 178 economies and for the Euro Area. Data concerning nominal exports and imports (in USD) and short-term money market interest rates (national interbank offered rates) are extracted from the on-line database IMF IFS. Data concerning real GDP per capita and nominal GDP (in USD) are extracted from the on-line database IMF World Economic Outlook (WEO) Database (September 2011).

3.2 Period and countries

This paper contains an analysis of 27 EU member countries (EU27) and selected subsamples named EU10 (the "new" member countries from Central and Eastern Europe, i.e. Bulgaria, the Czech Republic, Estonia, Hungary, Poland, Latvia, Lithuania, Romania, Slovakia, and Slovenia) and EU17 (the "old" member countries, i.e. Austria, Belgium, Cyprus, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Malta, the Netherlands, Portugal, Spain, Sweden, and the United Kingdom). The analysed time period is from 1993 to 2008. The

[3] See also Kucerova [10] for our earlier research focused solely on Central and Eastern European countries

reason for shortening the time span is the structural break from 2008 by reason of the world financial crisis. The data for the period 2009-2010 do not conform to the trend of the previous years due to a severe deviation from it by reason of the world financial crisis. Therefore, it is better to remove these data from the data sample.

3.3 Methods

The panel regression method (with fixed effects) is used. Panel data enable to survey a number of entities over several time periods. According to Baltagi [2], using panel data, we can check for the individual heterogeneity of the data and study the dynamics of adjustment. Panel data produce more informative data, more variability, less collinearity among the variables, more degrees of freedom and more efficiency. Panel data are able to identify and measure effects that cannot be detected in cross-section or time-series data. Because of the existence of reverse causality, i.e. the existence of simultaneity[4] (an important form of endogeneity) between the dependent variable and one of the explanatory variables of the "basic" model, the Instrumental Variables (IV) method and the Two Stage Least Squares Method (2SLS) is used (the "new" model). These methods are used for estimating the Simultaneous Equations Model (SEM). SEM applications with panel data are very powerful, because they allow us to check for unobserved heterogeneity while dealing with simultaneity (Wooldridge [16]).

4 Empirical model

4.1 The "basic" model

The basic specification of the empirical model is as follows:[5]

$$IFI_{it} = \beta_0 + \beta_1 TRADE_{it} + \beta_2 GDP_{it} + \beta_3 IBOR_{it} + \varepsilon_{it}, \quad (1)$$

where $TRADE_{it}$ is an explanatory variable which measures the trade openness of economy i and is defined as an arithmetic mean of exports and imports (i.e. the sum of exports plus imports divided by two) in country i in time t, GDP_{it} is the level of the real GDP per capita of country i in time t, and $IBOR_{it}$ is a short-term money market interest rate (the national interbank offered rate) of country i in time t. The dependent variable IFI_{it} is an indicator of financial integration. It is a *quantity-based measure* of financial integration and is constructed as an arithmetic mean of the stock of total foreign assets and the stock of total financial liabilities of country i in time t. Baltagi [2] states that by using panel data, one can also avoid the problem of spurious regression (in the case of non-stationary data). However, it is necessary to test for the stationarity. In the case of non-stationary data we have to transform the data to become stationary (by using the technique of first differencing). By using the ADF-Fisher Chi-square test we proved that the time series of the dependent variable (IFI) and some explanatory variables (GDP, $TRADE$) are non-stationary. Therefore, it is necessary to use the first differences of these variables:

$$\Delta(IFI_{it}) = \beta_0 + \beta_1 \Delta(TRADE_{it}) + \beta_2 \Delta(GDP_{it}) + \beta_3 IBOR_{it} + \varepsilon_{it} \quad (2)$$

		IFI	
	EU27	**EU10**	**EU17**
TRADE	2.66	0.84	4.05
	(1.63)*	(2.21)***	(1.48)*
GDP	4.85	1.59	5.02
	(2.10)**	(2.56)***	(1.23)
IBOR	234.36	-46.75	-2975.78
	(1.40)	(-0.77)	(-1.79)*
c	86163.22	3689.42	127912.4
	(4.85)***	(1.99)***	(3.89)***
No. of observations	375	129	246
Adjusted R-squared	0.43	0.53	0.39

Notes: Unstandardised coefficients. Heteroskedasticity-corrected *t*-statistics are shown in parentheses. *, **, *** indicate statistical significance at the 10 percent, 5 percent, and 1 percent confidence levels.

Table 1 Results of the model in equation (2) for EU27, EU10 and EU17, 1994-2008

[4] One or more of the explanatory variables is jointly determined with the dependent variable.
[5] See Lane and Milesi-Ferretti [13].

The results show that there is a positive relationship between the explained variable *IFI* and the explanatory variables *TRADE* and *GDP* in all analysed regions. The strongest impact of these variables can be found in the group of the EU17 countries. These results reflect the theoretical assumptions and empirical findings. The lower impact of the real GDP per capita on financial integration in the EU10 countries (compared to the EU17 countries) can be explained by an initial decrease of economic activity during the transformation process (from 1993 to 2000). However, results concerning the real GDP per capita in the EU17 countries are not statistically significant. The impact of the variable *IBOR* is ambiguous. There is a positive relationship between the *IFI* and *IBOR* variables, but only in the group of the EU27 countries and not in the two subsamples (EU10 and EU17). In these two subsamples, we can find a negative impact interest rates on the financial integration (especially in the EU17 countries, where these results are significant), which reflects the above mentioned theoretical assumptions.

4.2 The "new" model

Because of the existence of simultaneity between the dependent variable (*IFI*) and one of the explanatory variables (*TRADE*) of the "basic" model, the IV method and the 2SLS method are used in order to estimate the simultaneous equations model. The specification of the "new" model (the simultaneous equations model) is as follows:

$$\Delta(IFI_{it}) = \beta_0 + \beta_1\Delta(TRADE_{it}) + \beta_2\Delta(GDP_{it}) + \beta_3 IBOR_{it} + \varepsilon_{it},$$
$$\Delta(TRADE_{it}) = \beta_4 + \beta_5\Delta(IFI_{it}) + \beta_6\Delta(GDP_{it}) + \beta_7 EXCH\_RATE_{it} + \beta_8 IBOR_{it} + \varepsilon_{it} \quad (3)$$

where $EXCH\_RATE_{it}$ is the effective real exchange rate of country *i* in time *t*. The second equation (the trade function) satisfies the order condition because the *EXCH_RATE* variable is omitted from the first equation (the financial integration function), i.e. the *EXCH_RATE* variable appears only in the second equation. The rank condition for identifying the first equation is that the *EXCH_RATE* variable has a nonzero coefficient in the trade equation. The variables *GDP* and *IBOR* appear in both equations. The results of the estimation of the first stage (the trade function, i.e. the second equation in model (3)) of the simultaneous equations model are summarised in Table 2.

	TRADE		
	EU27	**EU10**	**EU17**
IFI	0.01	0.11	0.01
	(2.09)**	(4.80)***	(2.39)***
GDP	1.97	1.32	2.02
	(6.85)***	(6.13)***	(7.22)***
EXCH_RATE	43.29	79.20	-329.82
	(3.73)***	(10.88)***	(-4.35)***
IBOR	-35.96	-34.08	-7.98
	(-1.54)*	(-2.38)***	(-0.10)
c	3910.81	-5272.11	44195.85
	(2.86)***	(-6.65)***	(5.61)***
No. of observations	375	129	246
Adjusted R-squared	0.69	0.84	0.63

Notes: Unstandardised coefficients. Heteroskedasticity-corrected *t*-statistics are shown in parentheses. *, **, *** indicate statistical significance at the 10 percent, 5 percent, and 1 percent confidence levels.

Table 2 Results of the the first stage of the simultaneous equations model in equation (3) for EU27, EU10 and EU17, 1994-2008

The results of the trade function confirm that financial integration and economic growth have a positive impact on foreign trade. These findings correspond to the theoretical assumptions. Results concerning the effective real exchange rates are rather miscellaneous. Exchange rate appreciation leads to more intensive foreign trade especially in the EU10 and also EU27 countries, but this result does not reflect the theoretical assumptions. It can be ascribed to the convergence process of the EU10 countries which these countries underwent especially in the first half of the analysed time period. This process was accompanied by rising foreign trade as a result of finding new trade opportunities in developed European countries instead of former Soviet Union countries. There is also a strong and negative impact of money market interest rates on financial integration in all EU countries, i.e. higher interest rates may distort the foreign trade of these countries, especially of the EU10 countries. This result is quite important at the time of world economic crises when any attempt of commercial banks to increase interest rates on loans may seriously damage foreign trade by restricting trade finance products.

The results of the estimation of the second stage (the financial integration function, i.e. the first equation in model (3)) of the simultaneous equations model are summarised in Table 3. A positive relationship between the explained variable *IFI* and the explanatory variables *TRADE* and *GDP* is confirmed. The results are slightly different in this "new" model. The impact of the *TRADE* variable on the *IFI* variable is stronger (approximately three times higher). The same holds for the *GDP* variable, all analysed groups experienced a stronger effect of economic growth (measured by the real GDP per capita) on financial integration in this "new" model. The EU10 countries experienced the greatest (and most significant) effect of the GDP variable on financial integration in this "new" model; this finding can reflect a high importance of economic growth in the former transformation countries. But the results are not as highly significant as the results concerning the *TRADE* variable in the case of the EU27 and EU17 countries. The results of the *IBOR* variable are also different from the results presented in Table 1. A negative impact of the *IBOR* variable is proved in all three groups of countries but they are significant only in the case of the EU17 countries. However, a negative impact of money market interest rates on financial integration was measured in all cases, i.e. it is fully consistent with the theoretical assumptions. In other words, the rising money market interest rates hindered the process of financial integration in these countries, especially in the EU17 countries where the coefficient is much higher than in the EU10 countries.

		IFI	
	EU27	**EU10**	**EU17**
TRADE	8.94	1.83	13.41
	(2.24)**	(3.07)***	(2.42)***
GDP	4.92	3.23	7.20
	(0.69)	(3.15)***	(0.66)
IBOR	-353.84	-79.43	-11125.03
	(-0.67)	(-0.94)	(-2.20)**
c	126690.40	6607.024	229392.5
	(13.09)***	(6.04)***	(6.91)***
No. of observations	348	119	229
Adjusted R-squared	0.46	0.49	0.47

Notes: Unstandardised coefficients. Heteroskedasticity-corrected *t*-statistics are shown in parentheses. *, **, *** indicate statistical significance at the 10 percent, 5 percent, and 1 percent confidence levels.

Table 3 Results of the second stage of the simultaneous equations model in equation (4) for EU27, EU10 and EU17, 1994-2008

According to the results presented above, we can conclude that there is a high and significant relationship between financial and trade integration. Foreign trade raises financial linkages and greater financial cooperation and integration helps foreign trade to become more intensive. The processes of financial integration and trade integration are thus interconnected. By using the simultaneous equations model the results are stronger and more significant. To develop and improve the Euro Area, it is necessary to intensify European foreign trade to foster financial market integration.

5 The role of foreign trade and trade finance in the process of financial integration

The world financial crisis provoked national, international or supranational authorities to suddenly tighten financial regulation and supervision after a decade of loose regulation and supervision. However, these activities can seriously distort world trade on the ground of limiting trade finance. In January 2012, the Banking Commission of the International Chamber of Commerce together with the International Monetary Fund jointly published the *ICC-IMF Market Snapshot* containing the trade finance outlook for the year 2012 (International Chamber of Commerce [7]). The results of the survey are rather miscellaneous. Some 16 – 19% of the banks from advanced markets expect an improvement of the demand for trade finance during 2012, whilst this share in the case of ASEAN countries, China and India is 59%. On the other hand, *48% of the Euro Area banks anticipate a deterioration of the demand for trade financing in 2012*. It should be emphasised that **a large share of trade finance** (over a half of trade finance products offered by banks worldwide) *comes from the Euro Area banks* according to the survey results. These findings are really worrisome, especially in the light of the results presented in Chapter 4. Limiting trade finance can seriously damage the process of financial integration necessary for the successful functioning of the Euro Area (as an area with a single currency). Results also show, that some 74% of respondents stated that *the preparation for the implementation of Basel III had affected the costs of funds and liquidity for trade finance*. These findings seem to be quite worrisome, as it can decrease the available trade finance and thus hinder world trade. The European deleveraging process and limited fiscal possibilities of national govern-

ments could further undermine the economic recovery. Moreover, *integrated financial markets could cause another problem by the mutual dependence of integrated economies and thereby by the immediate transmission of a financial contagion from one country to another* no matter how distant they are.

6 Conclusion

This paper analysed the degree of financial integration of the EU countries with the rest of the world from 1993 to 2008 by using quantity-based measures of financial integration derived from the countries' international investment positions. The relations between the countries' foreign assets and liabilities and selected macroeconomic variables were investigated. Our results of the "basic" model confirmed a significant positive impact of foreign trade and the real GDP per capita on financial integration. The impact of money market interest rates on financial integration was ambiguous. Because of the existence of simultaneity between the dependent variable and one of the explanatory variables (foreign trade) of the "basic" model, the instrumental variables method and the two stage least squares method were used in order to estimate the simultaneous equations model with two functions (the financial integration function and the trade function). In the "new" model, the impact of trade was even stronger in all analysed regions (i.e. EU27, EU10 and EU17). The effect of economic growth was also a bit stronger in all three country groups (especially in the EU10 countries). The results concerning money market interest rates were more robust in the "new" model; a negative impact of money market interest rates on financial integration was measured in all cases. We also identified a positive impact of financial integration and the real GDP per capita on foreign trade in all regions. Real exchange rate appreciation led to more intensive foreign trade (especially in EU10), probably as a result of the transformation process in the EU10 countries associated with new trade opportunities. We concluded that foreign trade led to stronger financial linkages between countries and vice versa, i.e. that greater financial integration intensified the foreign trade linkages. However, world trade can be undermined by limiting trade finance as a result of tightening financial regulation and supervision after a decade of loose regulation and supervision (e.g. by the implementation of the Basel III capital framework). Thus, a stronger regulatory framework may hinder the world economic growth by limiting trade finance and thus by limiting world trade and financial integration. Therefore, tightening the financial regulation and supervision can be extremely harmful and can put the economic recovery behind.

References

[1] Baele, L., Ferrando, A., Hördahl, P., Krylova, E., and Monnet, C.: Measuring Financial Integration in the Euro Area. *European Central Bank Occasional Paper* **14** (2004).

[2] Baltagi, B. H.: *Econometric Analysis of Panel Data*. John Wiley & Sons Ltd., Chichester, 2005.

[3] Buch, C. M., Carstensen, K., and Schertler, A.: Macroeconomic Shocks and Foreign Bank Assets. *Kiel Institute for World Economics Working Paper* **1254** (2005).

[4] European Commission: European Financial Stability and Integration Report 2010. *Commission Staff Working Paper* **SEC(2011) 489** (2011).

[5] Goldberg, L. S.: The International Exposure of U.S. Banks. *National Bureau of Economic Research Working Paper* **11365** (2005).

[6] International Chamber of Commerce. ICC-IMF Market Snapshot. *A Market Intelligence Report of the ICC Banking Commission* **858E** (2012).

[7] Kose, M. A., Prasad, E. S., and Taylor, A. D.: Thresholds in the Process of International Financial Integration. *Global Economy and Development Working Paper* **35** (2009).

[8] Kotlanova, E., and Kotlan, I.: Vliv institucionálního prostředí na velikost korupce: empirická analýza. *Politická ekonomie* **2** (2011), 167-186.

[9] Kucerova, Z. Measuring Financial Integration in Central Europe through International Investment Positions. *Eastern European Economics* **4** (2009), 25-41.

[10] 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).

[11] Lane, P. R., and Milesi-Ferretti, G. M.: Long-Term Capital Movements. *International Monetary Fund Working Paper* **WP/01/107** (2001).

[12] Lane, P. R., and Milesi-Ferretti, G. M.: International Financial Integration. *International Monetary Fund Working Paper* **WP/03/86** (2003).

[13] Lane, P. R., and Milesi-Ferretti, G. M.: Capital flows to Central and Eastern Europe. *International Monetary Fund Working Paper* **WP/06/188** (2006).

[14] Lane, P. R., and Milesi-Ferretti, G. M.: The External Wealth of Nations Mark II. *Journal of International Economics* 73 (2007), 223-250.

[15] Wooldridge, J. M.: *Introductory Econometrics: A Modern Approach*. South-Western Cengage Learning, Mason, 2009.

Markets, Social Networks, and Endogenous Preferences[1]

Michal Kvasnička[2]

Abstract. This paper generalizes the Bell's model ("Locally interdependent preferences in a general equilibrium environment," JEBO, 2002), and models an interaction between a market, endogenous preferences, and a general social network. Contrary to Bell's results, 1) the system need not to converge, 2) the agents' preferences need not to be polarized, 3) the agents' preferences need not to adjust in the proportion to the availability (only the more abundant good can be consumed in one type of equilibrium), and 4) the agents with the same preferences need not to be clustered.

Keywords: endogenous preferences, market, social network, agent-based simulation

JEL classification: D83, R13, D51, D85, C69
AMS classification: 68U20, 05C82, 91D30, 91B24, 91B42, 91B52, 91B69

1 Introduction

In economic theory, agents' preferences are usually taken as exogenous and fixed. While there are many appealing reasons for this approach, it seems more than likely that preferences of real-world people are endogenous, change over time, and are influenced by their social interactions. There is a growing body of literature on endogenous preferences and impact of social networks on economic behavior. One of the studies in the preference adaptation in the context of social networks is A. M. Bell's seminal paper "Locally interdependent preferences in a general equilibrium environment" [1]. In this paper, Bell explored the evolution of agents' preferences by means of an agent-based simulation. In her basic model, the prices of two goods were set in a centralized market. The agents' preferences were endogenous: the agents increased their taste for the good that was consumed more in their local neighborhood. In equilibrium, each agent consumed only one of the goods, the agents with the same preferences were geographically clustered, and fewer agents specialized in consumption of the scarcer good.

Bell simulated her model only for one size and one type of social network (the grid). Therefore it is not known whether her findings generalize for other sizes and types of networks. It is also not known whether her findings were caused by the centralized market, the social network, or their interaction. The purpose of this study is to extend Bell's model to find answers to these two questions. To do it, the basic version of Bell's model (the model with a given endowment and no production) will be extended and simulated for various sizes and types of networks and their parametrization. It will be explored how much the characteristics of the resulting equilibrium depend on the characteristics of the social network. One special case discussed is *disconnected* agents. This will help us to understand what part of the Bell's findings are caused by the social network, and what are caused by the centralized market. It will be shown that many of Bell's results have to be modified.

2 Model

The used model is a straightforward generalization of the Bell's model of "exchange economy", see [1]. The only change is that the implicit grid network was replaced with the explicit description of a general social network. In general, the agents consume two kinds of goods (e.g. black and white t-shirts) which are comparable, i.e. their total consumption has a natural meaning (how many t-shirts an agent consumes). In every period, all agents get an initial endowment of each good. Then they trade these goods with each other at the centralized market at the market clearing price. The agents' preferences evolve over the time: each agent increases her preference for the good that has been recently more popular (i.e. more consumed) in her neighborhood. Her neighborhood consists of the agent herself and the agents she has relationship with. The set of all relationships in the population forms a social network.

[1]This paper has been created as a part of project of specific research no. MUNI/A/0796/2011 at the Masaryk University.
[2]Masaryk University, Faculty of Economics, Department of Economics, Lipová 41a, Brno, 602 00, the Czech Republic, michal.kvasnicka@econ.muni.cz

More formally, there are N agents indexed $i = 1, \ldots, N$. In every period, each agent gets the same endowment: e_1 units of good 1 and e_2 units of good 2 ($e_1 \leq e_2$). Agent i then demands x_{i1} unit of good 1 and x_{i2} units of good 2 to maximize her one-period Cobb-Douglas utility function subject to the constraint given by her endowment, i.e.

$$\max_{x_{i1}, x_{i2}} x_{i1}^{a_{it}} x_{i2}^{1-a_{it}} \qquad \text{s.t.} \quad p_1 x_{i1} + p_2 x_{i2} = p_1 e_1 + p_2 e_2, \tag{1}$$

where p_1 and p_2 are the prices of good 1 and good 2 respectively, and a_{it} is agent i's relative preference for good 1 at time t. The initial value of the preference parameter a_{i0} is drawn independently for each agent from the continuous uniform distribution $U(0, 1)$.

Agent i's demand for the two goods is then

$$x_{i1}(p_1, p_2) = a_{it}\left(e_1 + \frac{p_2}{p_1} e_2\right), \qquad x_{i2}(p_1, p_2) = (1 - a_{it})\frac{p_1}{p_2}\left(e_1 + \frac{p_2}{p_1} e_2\right). \tag{2}$$

Since the total endowment is given, the market clearing relative price is

$$\frac{p_1^*}{p_2^*} = \frac{e_2 \sum_{\forall j} a_{jt}}{e_1 \sum_{\forall j} (1 - a_{jt})} \tag{3}$$

The social network is represented by an undirected graph G, in which agents are vertices and their relationships are edges (connections). Two agents i and j have a relationship if they are connected with an edge; we then write $i \sim j \in G$. We define the agent i's neighborhood $n(i) = \{j : j = i \vee j \sim i \in G\}$, i.e. as the set of indices of all agents who are connected to agent i and the index of the agent i herself. The social network is for each simulation created randomly by a given algorithm; it is fixed within the simulation.

After observing the consumption in her neighborhood, each agent adjusts her preferences in such a way that she increases the preference for the good that is consumed more in her neighborhood. Specifically, agent i sets the future value of her preference parameter $a_{i,t+1}$ at

$$a_{i,t+1} = a_{it} + r\left(\frac{\sum_{j \in n(i)} x_{j1}}{\sum_{j \in n(i)} x_{j1} + \sum_{j \in n(i)} x_{j2}} - 0.5\right) \tag{4}$$

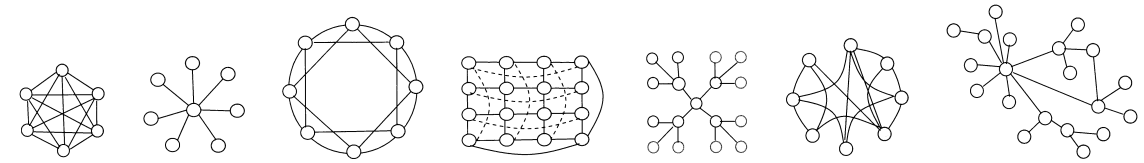
where the adjustment parameter $r \in (0, 1)$ regulates the speed of the preference adjustment.

The evolution of agents' preferences and consumption is simulated in the agent-based computational fashion (for introduction to it, see e.g. [3]). First, the model is initialized: a given social network consisting of N agents is created (see [5] for the network-creation algorithms), and each agent is assigned a random initial preference $a_{i,0}$ drawn from $U(0, 1)$. The simulation than proceeds in steps repeated until the model converges (i.e. the agents' preference parameters change no more), or the maximal amount of steps is reached. In each step, 1) the market clearing relative price p_1^*/p_2^* is calculated for the agents' current preferences (equation 3), 2) each agent's equilibrium consumption is calculated (equation 2), and 3) the agents' preferences are adjusted (equation 4).

3 Results of simulations

The model has been simulated for $N = 12, 25, 100, 156, 506, 992, 1980, 2500$ agents and the following standard types of network (see Figure 1 for the intuition or [5] for a detailed description of the networks): *grid* on torus, *star*, *ring*, *tree*, *small world*, *power* network, *complete* network and *disconnected* agents (for the *complete* network, the maximal N was 992 because the used simulation software was not able to carry more connections). In the ring network and initial small world network, each agent could have a connection with 2, 4, 6, or 8 closest agents. In the tree network, each agent could have 2 to 5 outgoing connections. In the power network, each agent was created with 1 to 5 new connections. The rewire probability in the small world network was $0.05, 0.1, 0.25, 0.5$. The endowment of good 1 was $e_1 = 1, 5, 10, 20, \ldots, 90, 95, 99, 100$, while the endowment of good 2 was $e_2 = 100$. The adjustment constant r was set to 0.5 as in [1]. The model has been simulated thirty times for each feasible combination of parameters. The maximal amount of simulation steps was set to 10 000. The total number of the simulation runs was 123 900. The model was simulated in NetLogo 5.0 [4]. The simulation results have been analyzed in R [2].

Since the full parametric space of the simulation is huge and non-rectangular, only the most salient stylized facts will be presented here. Specifically, the possible outcomes of the simulations, the role of the market and the interactions in the social network, clusters, and the determinants of the relative price will be discussed.



(a) complete (b) star (c) ring (d) grid (e) tree (f) small world (g) power

Figure 1 Examples of the network structures. The figures are taken from [5]. The grid is depicted as von Neumann neighborhood only for simplicity; Moore neighborhood was used in simulations.

3.1 Possible simulation outcomes

Bell [1] discusses only two kinds of simulation outcomes, both very similar. In both, the system converges, each good is consumed by some agents, and fewer agents consume the scarcer good than the more abundant one. The difference between the outcomes lies in the agents' equilibrium preferences. In the first case, each agent specializes in consumption of only one kind of good (i.e. $a_{it} \in \{0, 1\}$), while in the second case, at least some agents consume both goods (i.e. have $a_{it} \in (0, 1)$). Bell claims that the second outcome is unstable, and hence cannot occur in a simulation [1, p. 321]. Indeed, all Bell's simulation outcomes were of the first kind (she calls it "polarized").

My more general simulations produced a richer set of outcomes. There were four states: 1) the simulation did not converged (0.205 % of runs), 2) it converged but some agents remained non-polarized (0.220 % of runs), 3) the simulation converged, all agents were polarized but consumed only the more abundant good (13.835 % of runs), and 4) the simulation converged, all agents were polarized and both kinds of good were consumed by some agents which is the Bell's only outcome (85.739 % of runs). Since the above stated frequencies are somewhat artificial, it is instructive to decompose them.

Let us start with the state 1, the non convergence. This only occurred when the endowment of the two goods was precisely the same (i.e. $e_1 = e_2$), the network was either *complete* or *star*, and the number of agents was relatively large. In particular, the simulation of *complete* network always converged with $N = 12$ and in 93.333 % of runs with $N = 25$; otherwise it did not converge. The simulation of *star* network always converged with $N = 12, \ldots, 100$; then the frequency of the converged runs decreased with the number of agents: 93.333 % of runs with $N = 256$ converged, 70 % of runs with $N = 506$, 40 % of runs with $N = 992$, 33.333 % of runs with $N = 1980$, and 10 % of runs with $N = 2500$. It is interesting that these two types of network are in a sense opposite to each other and the reasons why they do not converge are opposite too. In the *complete* network, all agents are symmetric. If the network does not converge, it is because each agent interacts directly with all other agents and there is no "locality". Thus any change of an agent's preferences forces all other agents to change their preferences, which makes the next adjustment necessary, and so on. In theory, there can be both many polarized and non-polarized equilibria (e.g. one half agents prefer good 1 and the other half prefer good 2, or each agent's preference is 0.5 etc.) but all these equilibria are very fragile. In the *star* network, one central agent is connected to all other (branch) agents, while each branch agent is connected only to the central agent. The branch agents tend to be polarized in this case—roughly one half of them prefers good 1 and the other half good 2. It is the central agent who wavers between the two and thus precludes the convergence.

The state 2, converged but not-polarized, occurred most often when the endowment of the two kinds of good was the same too—96.703 % of runs in this state occurred when $e_1 = e_2$. However, it occurred with other proportions of the endowment too, even though only rarely. Most often this state occurred when $e_1 = e_2$, the network was either *complete* or *star*, and number of agents was small. All runs in the *complete* network with $N = 12$ and 6.667 % of runs with $N = 25$ ended in this state. (No other run on the *complete* network converged, see above.) As for the *star* network, all runs with $N = 12, \ldots, 100$, 93.333 % of runs with $N = 256$, 70 % of runs with $N = 506$, 40 % of runs with $N = 992$, 33.333 % of runs with $N = 1980$ and 10 % of runs with $N = 2500$ ended in this state. Beside these, two other networks (*power* and *small world*) tended to end up in this state. With few agents and many connections, these networks are close to the *complete* network. Usually, the *power* network ended in this state when $e_1 = e_2$ and $N = 12$. The frequency of this state raised with the number of connections from 3.333 % with 2 connections to 46.667 % with 5 connections. The *power* network with 5 connections ended in this state once also with $N = 25$ and once with $N = 256$ and $e_1 = 0.8e_2$. The *small world* network ended in this state most often with $N = 12$ and $e_1 = e_2$; however, it ended in this state several times even with other configuration, e.g. with $N = 1980$ and $e_1 = 0.5e_2$. Beside these, only *tree* network ended in this state once with $N = 25$ and $e_1 = 0.5e_2$. In general, the results for this state suggest that even though the non-polarized state is fragile, it can arise from the simulation, especially if the number of agents is small. It is also possible, that the runs

denoted as "non-convergent" might have converged if the simulation was given more time.

The state 3, the converged polarized state where only the more abundant good 2 is consumed, occurred with any network including the *disconnected* one. However, with *disconnected* network, this state occurred only when the amounts of the two goods were extremely asymmetric ($e_1 = 0.01e_2$ or $e_1 = 0.05e_2$) and the number of agents was small ($N = 12, \ldots, 100$ when $e_1 = 0.01e_2$ and $N = 12$ in the other case). With increase in the endowment of good 1 or the number of agents, the frequency of state 3 decreases and soon vanished with the *disconnected* network. The situation was dramatically different with *complete* network. With it, the state 3 occurred always when $e_1 < e_2$ no matter what was the number of agents. The behavior of other networks was between these two extremes. In general, the increase in the endowment of the scarcer good e_1 or in the number of agents decreased the probability that all agents consumed only good 2. However, individual networks somewhat differed in their propensity to this state. Paradoxically, *grid* network used by Bell was relatively more prone to end up in this state than most other networks, with exception of *complete* and *star* networks and *small world* network with the comparable number of connections. Surprisingly, the closest to the behavior of *complete* network was again *star* network. The overall probability that the state 3 occurs can be modeled with the logistic regression. The results are summarized in Table 1. The presence of a social network increases the probability that only the more abundant good 2 is consumed (all network dummy variables coefficients are positive). The rise in the number of agents or in the amount of the scarcer good decreases the probability, while the rise of the average number of connections between agents increases it.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−1.5718	0.1233	−12.75	0.0000
grid network	3.3075	0.1363	24.26	0.0000
power network	2.3564	0.1263	18.66	0.0000
small-world network	2.4546	0.1241	19.78	0.0000
ring network	2.6996	0.1269	21.28	0.0000
star network	4.5173	0.1359	33.23	0.0000
tree network	1.3319	0.1290	10.33	0.0000
complete network	10.4102	0.1689	61.62	0.0000
number of agents	−0.0013	0.0000	−65.32	0.0000
endowment e_1	−0.0796	0.0007	−108.29	0.0000
average number of connections	0.0018	0.0002	7.13	0.0000
log likelihood: −25089.08,	McFadden's pseudo R^2: 0.496,		correctly predicted: 92.65 %	

Table 1 Binary logit model. Dependent variable is one if the state is state 3—converged, polarized, and only the more abundant good is consumed; otherwise it is zero. Network dummy variables are in contrast to *disconnected* agents. The average number of connections in the tree network is only approximate. Only data from the converged runs were used.

3.2 Role of market and of social network

Bell claims that "the number of agents consuming a good in the steady state is proportional to the availability of the goods…" [1, p. 311]. If this was true, the relative price p_1/p_2 would be close to unity with no regard to the proportion of the endowments as long as $e_1 > 0$ and $e_2 > 0$. Bell does not attempt to judge what part of this result is caused by the centralized market and what part is caused by the interactions on the social network. She only claims that "the price acts as a negative feedback mechanism that limits consumptions of scarce … goods and encourages consumption of plentiful … goods" [1, p. 311]. Similarly, the "bandwagon" effect of the preference adjustment in the social network creates a positive feedback mechanism: if one good is more abundant than the other one, the agents can see it more often, and learn to like it more. (The previous section shows that the bandwagon effect can be so strong that all agents learn to consume only the abundant good, and throw away the scarce one, which contradicts the Bell's claim stated above.)

Disconnected agents were included among the networks to allow to disentangle the impact of the two types of feedback: *disconnected* "network" includes only the market negative feedback; the rest networks include both the negative feedback and the bandwagon positive feedback. Their impacts can be seen either in the relative

price p_1/p_2 or in the relative consumption, i.e. the average consumption of the agents who consume only good 1 divided by the average consumption of the agents who consume only good 2. However, the later is not necessary since it is the reciprocal value of the former. It is because all agents have the same endowment e_1 and e_2, and hence the same income $p_1e_2 + p_2e_2$. (The relative price $p_1/p_2 = 0$ means that no agent consumes the scarcer good 1, and hence the relative consumptions is not defined in this case. For this reason, the relative price p_1/p_2 will be used in the analysis.)

It is the market negative feedback (together with the preference-adjustment algorithm) what brings the system into an equilibrium (*disconnected* "network" always converges). In the equilibrium, usually both goods are consumed (the state 3 occurs least often with *disconnected* agents). And it is this force that presses the relative price p_1/p_2 to unity. The left part of Table 2 shows that the scarcer good 1 is usually more expensive than the more abundant good 2 ($p_1/p_2 \geq 0$) when agents are disconnected but as e_1/e_2 and the number of agents raises, the market negative feedback is able to push the relative price p_1/p_2 to unity. The consumers of the scarce good 1 have in average lower consumption that the consumers of the abundant good 2 in this case.

The presence of a social network adds the bandwagon effect which destabilizes the network (and possibly creates the clusters, see below). First, the bandwagon effect can preclude the convergence and increases the probability that the non-polarized state 2 occurs when the system converges (see above). Moreover, it raises the demand for the more abundant good and thus increases its price. The right part of Table 2 shows the example of the relative price for the *grid* network: it is almost always below unity, which means that the bandwagon positive feedback usually more than offsets the stabilizing impact of the market negative feedback. In the limit case, the scarcer good 1 is not consumed at all and its relative price $p_1/p_2 = 0$. If the scarce good 1 is consumed at all, its consumers have higher consumption than the consumers of the abundant good 2 in this case.

	disconnected							grid						
N / e_1	1	5	10	20	95	99	100	1	5	10	20	95	99	100
12	1.82	1.60	1.57	1.56	1.16	1.14	1.15	0.00	0.00	0.00	0.00	1.05	1.01	1.00
25	0.98	1.52	1.66	1.60	1.15	1.16	1.15	0.00	0.00	0.00	0.00	0.98	1.06	1.05
49	1.40	1.64	1.85	1.62	1.05	1.04	1.03	0.00	0.00	0.00	0.48	0.96	0.99	1.02
100	1.67	1.70	1.71	1.55	1.03	1.01	1.00	0.00	0.00	0.46	0.57	0.97	1.00	1.01
256	1.40	1.64	1.74	1.53	1.00	1.00	1.00	0.00	0.48	0.58	0.61	0.95	0.97	0.97
506	1.34	1.77	1.75	1.58	1.01	1.00	1.00	0.00	0.59	0.59	0.62	0.96	0.98	0.99
992	1.29	1.74	1.73	1.55	1.01	1.00	1.00	0.00	0.59	0.62	0.63	0.96	1.01	1.01
1980	1.33	1.77	1.73	1.56	1.01	1.00	1.00	0.00	0.60	0.61	0.66	0.94	0.97	0.97
2500	1.36	1.79	1.75	1.56	1.01	1.00	1.00	0.00	0.59	0.60	0.66	0.95	0.99	0.99

Table 2 Average relative price p_1/p_2 for selected endowments e_1 (columns), number of agents N (rows), and networks (panels). Averaging can bias the relative price slightly upward. Only data from states 3 and 4 were used in the computation.

3.3 Clusters

Bell reported that the agents with the same preference were clustered together in the polarized equilibria. This is indeed what happens in the polarized equilibria in the networks in which each agent has many connections (e.g. *grid* network). Each agent can retain her preference only if the number of the agents with the same preference is in her neighborhood higher than the number of agents with the opposite preference, i.e. if the agents are clustered. This, however, is not necessary in networks where some agents have only one connection (e.g. in *tree* network). There might survive also agents who have preference different from their surroundings. The sufficient condition is that 1) the dissenting agent i has only one connection, prefers the cheaper scare good 1 ($p_1/p_2 < 1$), and hence has higher consumption than the consumers of the more abundant good 2, and 2) agent j in the neighborhood $n(i)$ of agent i has more than one connection and she and most other agents in her neighborhood $n(j)$ consume the expensive abundant good 2, and hence have lower consumption than agent i. In this case, there can survive consumers of the scarce good 1 outside clusters; there even need not to be any clusters at all. This also explains relatively higher resistance of this type of networks against the state 3.

3.4 Quantitative determinants of relative price

The type and size of the social network do not affect only the qualitative, but also the quantitative characteristics of the equilibrium. Table 3 shows the determinants of the relative price p_1/p_2. The presence of the social network lowers the relative price. The rise in the number of agents and the endowment e_1 of the scarce good 1 brings the relative price to unity (notice that the corresponding parameters for the social network and *disconnected* agents have the opposite signs). Obviously, the used type of network strongly affects the equilibrium relative price.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.63	0.03	54.28	0.00
grid network	-1.37	0.03	-45.23	0.00
power network	-1.44	0.03	-47.89	0.00
small-world network	-1.35	0.03	-44.84	0.00
ring network	-1.27	0.03	-42.01	0.00
star network	-1.67	0.03	-54.71	0.00
tree network	-1.24	0.03	-41.30	0.00
complete network	-1.98	0.03	-64.68	0.00
number of agents	-0.00	0.00	-2.02	0.04
endowment e_1	-0.01	0.00	-18.85	0.00
average number of connections	-0.00	0.00	-5.04	0.00
number of agents × social network	0.00	0.00	6.01	0.00
endowment e_1 × social network	0.01	0.00	41.00	0.00
s: 0.2422, R^2: 0.6145, adjusted R^2: 0.6144, F: 1.6380 on 12 and 123360 DF, $p < 2.2e-16$				

Table 3 Linear regression model. Dependent variable is the relative price p_1/p_2. The network dummy variables and their summary "social network" are reported in contrast to *disconnected* agents. The robust standard errors of the parameter estimates are reported. The average number of connections in the tree network is only approximate.

4 Conclusions

The simulation results show that Bell's conclusions have to be somewhat modified. There are more possible outcomes than she expected: the system need not to converge, the agents' preferences need not to be polarized, and the number of agents consuming a good need not to be "proportional to the availability"—the agents can consume only the abundant good and throw away the scarce one. The agents need not to be clustered. In general, the convergence to an equilibrium, the type of the resulting equilibrium, and the equilibrium relative price are quite sensitive to the type and size of the used network because it is the bandwagon effect what destabilizes the system. This means that a careful specification of the social network might be crucial for modeling real-world markets.

References

[1] Bell, A. M.: Locally interdependent preferences in a general equilibrium environment, *Journal of Economic Behavior & Organization* **47** (2002), 309–333.

[2] R Development Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Wien, 2011, http://www.R-project.org/.

[3] Tesfatsion, L.: Agent-Based Computational Economics: A Constructive Approach to Economic Theory. In: *Handbook of Computational Economics*, vol. 2, (L. Tesfatsion and K. L. Judd, eds.), Elsevier, 2006, 831–880.

[4] Wilensky, U. (1999). *NetLogo*. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL, http://ccl.northwestern.edu/netlogo/.

[5] Wilhite, A.: Economic activity on fixed networks. In: *Handbook of Computational Economics*, vol. 2, (L. Tesfatsion and K. L. Judd, eds.), Elsevier, 2006, 1013–1045.

Bootstrap application of the Bornhuetter-Ferguson method

Bohdan Linda[1], Jana Kubanová[2]

Abstract. The insurance companies use many different methods for calculation of the claim reserves, like the chain-ladder method, the Bornhuetter-Ferguson method, the Poisson method and many others. Majority of these models were originally created as deterministic models. Currently, these models are modified as stochastic models. When the stochastic model provides the relevant results, we have to dispose of information about the probability distribution of processed data. Due to real situation that the insurance companies do not publish the actual data, major complications arise both with coefficients estimates in the models and with determination of their properties (bias, standard deviation etc.). The bootstrap principle is used to estimate these coefficients in the paper; the described process does not require knowledge of the probability distribution of processed data, which can be considered as an advantage at solution of this problem.

Keywords: bootstrap principle, cumulative loses, expected ultimate loses, premium, Bornhuetter-Ferguson principle, insurance benefit, future loses.

JEL Classification: C63**,** G17
AMS Classification: 62G05, 62M20

1 Introduction

The common situation in insurance practice is that the claim is not solved in the accident year but the insurance benefit is distributed in some following years. These events are divided into two basic groups that are connected with two types of claim reserves for past exposures, see Pacáková [6]:
a) IBNR (Incurred but not reported). The reserve for insurance benefit from the claims that have occurred but haven't been reported yet corresponds with it.
b) IBNS (Reported but not settled). This means not settled insured accident corresponding with the reserve for the insurance benefit from the claims that have been reported but have not been settled. The payment is expected in the future.

The insurance company has to create the adequate claim reserves for such situations. The table 1 describes the situation when we assume that the payment of the insurance benefit is distributed in n +1 years (development http://mme2012.opf.slu.cz/ .year).

accident	**development year *j***						
year *i*	**0**	**1**	**2**	**...**	***n*-2**	***n*-1**	***n***
0	$X_{0,0}$	$X_{0,1}$	$X_{0,2}$	**...**	$X_{0,n-2}$	$X_{0,n-1}$	$X_{0,n}$
1	$X_{1,0}$	$X_{1,1}$	$X_{1,2}$	**...**	$X_{1,n-2}$	$X_{1,n-1}$	$X_{1,n-1}$
2	$X_{2,0}$	$X_{2,1}$	$X_{2,2}$	**...**	$X_{2,n-2}$	$X_{2,n-1}$	$X_{2,n}$
...	**...**	**...**	**...**	**...**	…	…	…
***n* - 1**	$X_{n-1,0}$	$X_{n-1,1}$	$X_{n-1,2}$	**...**	$X_{n-1,n-2}$	$X_{n-1,n-1}$	$X_{n-1,n}$
n	$X_{n,0}$	$X_{n,2}$	$X_{n,2}$	**...**	$X_{n,n-2}$	$X_{n,n-1}$	$X_{n,n}$

Table 1 Insurance benefits

The first column of the table presents the year of the insurance event origin (accident year). Insurance benefits $X_{i,j}$ can be expressed as incremental values $Y_{i,j}$, or cumulative values $C_{i,j}$. The incremental insurance benefits $Y_{i,j}$ relating to the events that happened in the accident years $i = 0,1,2,\ldots,n$, are settled in the years $j = 0,1,2,\ldots,n$ (development year j). The cumulative insurance benefits $C_{i,j}$ present the sum of all insurance benefits in the de-

[1] University of Pardubice/Faculty of economics and administration, Department of mathematics, Studentská 95, 53210 Pardubice, bohdan.linda@upce.cz.
[2] University of Pardubice/Faculty of economics and administration, Department of mathematics, Studentská 95, 53210 Pardubice, jana.kubanova@upce.cz.

velopment years 0 till *j*. The relation between incremental and cumulative insurance benefits has the following form $C_{i,j} = \sum_{k=0}^{j} Y_{i,k}$.

The insurance company knows the values lying over the main diagonal (the white cells). The main problem, the insurance company states ahead is to estimate the values lying in the cells below the main diagonal. These estimates are called in case of $Y_{i,j}$ the incremental claim reserves, in case of $C_{i,j}$ the cumulative claim reserves. The grey colour marks these cells in table 1.

Many methods deal with solution of this problem. Some of them are described for example in Bornhuetter, and Ferguson [1], Fecenko [2], Mack [4], Pacáková [5], Stehlík et al. [7], Wuttrich, and Merz [10]. The methods like loss-development method, chain ladder method, Cape-Cod method, additive method and others are classified as a special version of the Bornhuetter-Ferguson principle in the works made by Schmidt [8] and Schmidt, Zocher [9].

2 Bornhuetter – Ferguson method as general principle

Variables $Y_{i,j}$ and $C_{i,j}$ are considered to be random variables with an unknown distribution. General Bornhuetter – Ferguson principle assumes the existence of the parameters $\alpha_1, \alpha_2, \ldots, \alpha_n$ and $\gamma_1, \gamma_2, \ldots, \gamma_n$ and it holds-true (Schmidt and Zocher [9])

$$E(C_{i,j}) = \alpha_i \gamma_j \qquad i, j \in \{1, 2, ..., n\} \tag{1}$$

where

$$\alpha_i = E(C_{i,n}) \quad \text{and} \quad \gamma_j = \frac{E(C_{i,j})}{E(C_{i,n})} \tag{2}$$

The unknown values $C_{i,j}$ are then predicted by a predictor

$$\hat{C}_{i,j} = \hat{\alpha}_i \, \hat{\gamma}_j \qquad i + j > n$$

where $\hat{\alpha}_i$ and $\hat{\gamma}_j$ are so called prior estimators of the parameters α_i and γ_j. These estimates can be obtained by different methods both from run-off triangle (upper triangle in the table 1), and from insurance market investigation. The difference $R_i = C_{i,n} - C_{i,n-i}$ is the accident year reserve and for its expected value can be written $E(R_i) = E(C_{i,n}) - E(C_{i,n-i})$. With regards to (2) this formula can be written in the form $E(R_i) = \alpha_i(1 - \gamma_{n-i})$. Then we use the estimator $\hat{R}_i$ for the estimate of the accident year reserve R_i

$$\hat{R}_i = \hat{\alpha}_i(1 - \hat{\gamma}_{n-i}) . \tag{3}$$

Schmidt [8] states that in the original Bornhuetter-Ferguson method for these coefficients holds true:

$$\hat{\alpha}_i = \pi_i \hat{\kappa}_i \quad \text{and} \quad \hat{\gamma}_{n-i} = \hat{\gamma}_{n-i}^{CL} ,$$

where π_i is premium income of accident year *i* and $\hat{\kappa}_i$ is estimate of the expected loss ratio $\kappa_j = E\,[\, C_{i,n} / \pi_i \,]$ of the accident year *i*. $\hat{\gamma}_{n-i}^{CL}$ is the chain ladder estimator of parameter γ_{n-i} .For the cumulative losses is true $C_{i,k} = C_{i,n-i} + (C_{i,k} - C_{i,n-i})$. Similarly to the formula (3) when we express the expected value $E(C_{i,k} - C_{i,n-i})$ by the help of parameters α_i and γ_k we get the formula $E(C_{i,k} - C_{i,n-i}) = \alpha_i(\gamma_k - \gamma_{n-i})$. On the base of this formula Schmidt, and Zocher [9] define the general Bornhuetter-Ferguson predictor of future cumulative losses for $i+k > n$ as

$$\hat{C}_{i,k}^{BF} = C_{i,n-i} + \hat{\alpha}_i(\hat{\gamma}_k - \hat{\gamma}_{n-i}) \tag{4}$$

The general Bornhuetter - Ferguson predictor says that the future cumulative insurance benefits are compounded from the part of already paid insurance benefits and from the estimate of the insurance benefits in the development years n-i+1 till k.

Schmidt [8] shows that the different methods of the claim reserves estimates (loss-development method, chain-ladder method, Cape-Cod method and additive method) are basically the different variants of the general Bornhuetter - Ferguson principle (4). They differ only by the construction of the estimators $\hat{\alpha}_i$ and $\hat{\gamma}_j$.

3 Application of the bootstrap in special versions of Bornhuetter – Ferguson principle

Wuthrich, and Merz [10] state in their book the different methods of the claim reserves estimate and the different methods of calculation of their parameters.

The bootstrap approach to the parameters estimates is implied in this book. The more detailed description of this approach of the claim reserves estimate by Chain ladder method is in Linda, Kubanová, and Jindrová [3].

The succeeding methods, which are presented as the special versions of the general Bornhuetter-Ferguson principle, use the estimates of the development coefficients $\hat{\lambda}_j$, obtained by chain-ladder method. That is why the shortened description of this estimate by bootstrap method is stated.

The starting point is the AR model of the cumulative losses

$$C_{i,j+1} = \lambda_j \cdot C_{i,j} + \sigma_j \sqrt{C_{i,j}}\, \varepsilon_{i,j+1},$$

where $\varepsilon_{i,j+1}$ present a random residuals. The bootstrap replication $C^*_{i,j+1}$ for the upper triangle of the table 2 were gained by the help of resampling of the adjusted residuals

$$z_{i,j+1} = \hat{\varepsilon}_{i,j+1}\left(1 - C_{i,j}\left(\sum_{i=0}^{n-j-1} C_{i,j}\right)^{-1}\right)^{-\frac{1}{2}},$$

where $\hat{\varepsilon}_{i,j+1} = \hat{\sigma}_j^{-1}\left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{\lambda}_j^{CL}\right)\sqrt{C_{i,j}}$ and $\hat{\lambda}_j^{CL} = \left(\sum_{i=0}^{n-j-1} C_{i,j+1}\right)\left(\sum_{i=0}^{n-j-1} C_{i,j}\right)^{-1}$. The bootstrap replications of the estimate of the development coefficient were calculated according to the formula

$$\hat{\lambda}_j^{CL*} = \left(\sum_{i=0}^{n-j-1} C^*_{i,j+1}\right)\left(\sum_{i=1}^{I-j} C^*_{i,j}\right)^{-1}, \text{ where } C^*_{i,j+1} = \hat{\lambda}_j C^*_{i,j} + \hat{\sigma}_j \sqrt{C^*_{i,j}}\, z^*_{i,j+1}.$$

When the chain ladder process in the general Bornhuetter – Ferguson model $\hat{C}^{BF}_{i,k} = C_{i,n-i} + \hat{\alpha}_i(\hat{\gamma}_k - \hat{\gamma}_{n-i})$ is used, then the bootstrap replication of the parameter $\hat{\gamma}_k$ is calculated according to the formula

$$\hat{\gamma}^{*CL}_k = \prod_{j=k+1}^{n} \frac{1}{\hat{\lambda}^{*CL}_k}. \tag{5}$$

Three different versions of the Bornhuetter – Ferguson method are presented in this paper.

3.1 Loss-development method

Loss development method estimator of future cumulative losses is given by formula

$$\hat{C}^{LD}_{i,k} = \hat{\gamma}_k \frac{C_{i,n-i}}{\hat{\gamma}_{n-i}}.$$

According to Schmidt [8] this formula can be expressed as a special version of the Bornhuetter – Ferguson principle

$$\hat{C}^{LD}_{i,k} = C_{i,n-i} + \frac{C_{i,n-i}}{\hat{\gamma}_{n-i}}(\hat{\gamma}_k - \hat{\gamma}_{n-i}) \qquad \text{where} \qquad \alpha_i = \frac{C_{i,n-i}}{\hat{\gamma}_{n-i}}.$$

Bootstrap replication of future cumulative losses are calculated according this formula as

$$\hat{C}^{*LD}_{i,k} = C_{i,n-i} + \frac{C_{i,n-i}}{\hat{\gamma}^*_{n-i}}(\hat{\gamma}^*_k - \hat{\gamma}^*_{n-i}) \tag{6}$$

where $\hat{\gamma}^*_i$ are given by formula (5).

3.2 Cape-Cod method

According to Schmidt [8] the Cape-Cod method predictions of the future cumulative losses are given by the formula

$$\hat{C}_{i,k}^{CC} = C_{i,n-i} + \pi_i \hat{\kappa}^{CC} (\hat{\gamma}_k - \hat{\gamma}_{n-i})$$

which is again the special version of the Bornhuetter-Ferguson principle, where $\hat{\alpha}_i = \pi_i \hat{\kappa}^{CC}$. The coefficient $\hat{\kappa}^{CC}$ is called the Cape-Cod loss ratio and it can be expressed by the formula

$$\hat{\kappa}^{CC} = \frac{\sum_{j=0}^{n} C_{j,n-j}}{\sum_{j=0}^{n} \hat{\gamma}_{n-j} \pi_j},$$

where π_i is the premium of the accident year.

When we use bootstrap for calculation of the future losses, we have to calculate at first the bootstrap replications of the Cape-Cod ratio $\hat{\kappa}^{*CC} = \frac{\sum_{j=0}^{n} C_{j,n-j}}{\sum_{j=0}^{n} \hat{\gamma}_{n-j}^{*} \pi_j}$, where $\hat{\gamma}_i^{*}$ are given by formula (5) again. Than the bootstrap predictor of the Cape Code method is

$$\hat{C}_{i,k}^{*CC} = C_{i,n-i} + \pi_i \hat{\kappa}^{*CC} (\hat{\gamma}_k^{*} - \hat{\gamma}_{n-i}^{*})$$

3.3 Additive method

The additive method predictors of the future cumulative losses have expression

$$\hat{C}_{i,k}^{AD} = C_{i,n-i} + \pi_i \sum_{j=n-i+1}^{k} \hat{\varsigma}_j^{AD}$$

where $\hat{\varsigma}_k^{AD} = \left(\sum_{j=0}^{n-k} X_{j,k} \right) \left(\sum_{j=0}^{n-k} \pi_j \right)^{-1}$. Schmidt [8] shows, that this model can be expressed in the Bornhuetter-Ferguson form as $\hat{C}_{i,k}^{AD} = C_{i,n-i} + \hat{\alpha}_i^{AD} (\hat{\gamma}_k^{AD} - \hat{\gamma}_{n-i}^{AD})$, where $\hat{\alpha}_i^{AD} = \pi_i \sum_{j=0}^{n} \hat{\varsigma}_j^{AD}$ and $\hat{\gamma}_k^{AD} = \left(\sum_{j=0}^{k} \hat{\varsigma}_j^{AD} \right) \left(\sum_{j=0}^{n} \hat{\varsigma}_j^{AD} \right)^{-1}$.

Bootstrap Additive method predictors are then

$$\hat{C}_{i,k}^{AD*} = C_{i,n-i} + \hat{\alpha}_i^{AD*} (\hat{\gamma}_k^{AD*} - \hat{\gamma}_{n-i}^{AD*})$$

where $\hat{\alpha}_i^{AD*} = \pi_i \sum_{j=0}^{n} \varsigma_j^{AD*}$, $\hat{\gamma}_k^{AD*} = \left(\sum_{j=0}^{k} \varsigma_j^{AD*} \right) \left(\sum_{j=0}^{n} \varsigma_j^{AD*} \right)^{-1}$ and $\hat{\varsigma}_k^{AD*} = \left(\sum_{j=0}^{n-k} X_{j,k}^{*} \right) \left(\sum_{j=0}^{n-k} \pi_j \right)^{-1}$.

$X_{j,k}^{*}$ are the bootstrap replications of the incremental losses, that were gained by the analogical process as the replications $C_{j,k}^{*}$.

4 Application of the Bornhuetter – Ferguson bootstrap principle

To demonstrate the Bornhuetter – Ferguson bootstrap principle in a concrete way, we used data published by Pacáková [5], that are shown in the following table 2. We can see in the tables 2 till 5 results of individual methods. The results of the common chain ladder method are presented in the table 2 (the lower dark triangle) and this table serves for comparison the results with results obtained by Bornhuetter – Fergusson bootstrap method. Result of bootstrap loss-development method (table 3), bootstrap Cape-Cod method (table 4) and bootstrap additive method (table 5), all after 1000 bootstrap simulation, are stated in this article.

Values of the predictors $\hat{\gamma}_k^{CL}$ and $\hat{\gamma}_k^{AD}$, that were calculated by classical chain-ladder and additive method, and the bootstrap replications $\hat{\gamma}_k^{*CL}$ of the parameter γ_k^{CL} are compared in the table 6. The bootstrap method enables to calculate bias of the parameter estimator of interest; it is stated in the last column of the table 6. The last table 7 summarizes the values of the predictor "the expected ultimate loss", they are calculated by classical chain-ladder and additive method (the first two columns) and by bootstrap method (the third and fourth column). We can see in all tables, that the results of individual methods are not fundamentally different; the bootstrap method enables us to avoid the assumptions that are necessary for application of the classical methods.

		development year *j*					
	i	**0**	**1**	**2**	**3**	**4**	**5**
accident year *i*	**0**	566	1 049	1 270	1 407	1 460	1 483
	1	501	993	1 186	1 345	1 409	1 431
	2	543	1 055	1 287	1 471	1 534	1 535
	3	652	1 323	1 633	1 842	1 921	1 922
	4	739	1 479	1 799	2 030	2 116	2 117
	5	752	1 478	1 479	1 669	1 740	1 741
		$\hat{\lambda}_0$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_4$	
		1.966	**1.216**	**1.128**	**1.043**	**1.016**	

Table 2 Cumulative values of claim reserves

		development year *j*						$\hat{\alpha}_i^{LD}$
	i	**0**	**1**	**2**	**3**	**4**	**5**	
accident year *i*	**0**	566	1 049	1 270	1 407	1 460	1 483	1483
	1	501	993	1 186	1 345	1 409	1 431	1431
	2	543	1 055	1 287	1 471	1 534	1 534	1557
	3	652	1 323	1 633	1 837	1 921	1 915	1944
	4	739	1 479	1 794	2 102	2 116	2 103	2135
	5	752	1 469	1 470	1 722	1 740	1 723	2120

Table 3 Bootstrap loss-development method

		development year *j*						$\hat{\alpha}_i^{CC}$	π_i
	i	**0**	**1**	**2**	**3**	**4**	**5**		
accident year *i*	**0**	566	1 049	1 270	1 407	1 460	1 483	1453	1 700
	1	501	993	1 186	1 345	1 409	1 431	1436	1 680
	2	543	1 055	1 287	1 471	1 532	1 556	1538	1 800
	3	652	1 323	1 633	1 830	1 901	1 934	1880	2 200
	4	739	1 479	1 782	1 996	2 078	2 109	2051	2 400
	5	752	1 330	1 582	1 761	1 826	1 855	1709	2 000

Table 4 Bootstrap Cape-cod method

		development year *j*						$\hat{\alpha}_i^{LD}$	π_i
	i	**0**	**1**	**2**	**3**	**4**	**5**		
accident year *i*	**0**	566	1 049	1 270	1 407	1 460	1 483	1505	1 700
	1	501	993	1 186	1 345	1 409	1 432	1487	1 680
	2	543	1 055	1 287	1 471	1 533	1 558	1593	1 800
	3	652	1 323	1 633	1 837	1 913	1 943	1948	2 200
	4	739	1 479	1 790	2012	2 095	2 128	2125	2 400
	5	752	1 345	1 604	1 789	1 858	1 885	1771	2 000

Table 5 Bootstrap additive method

$\hat{\gamma}_k^{CL}$	$\hat{\gamma}_k^{AD}$	$\hat{\gamma}_k^{*CL}$	bias $\hat{\gamma}_k^{CL}$
0.350	0.360	0.355	0.005
0.688	0.695	0.693	0.004
0.837	0.841	0.840	0.003
0.944	0.946	0.945	0.001
0.984	0.985	0.984	0.000
1.000	1.000	1.000	0.000

Table 6 The calculated $\hat{\gamma}_k^{CL}$, $\hat{\gamma}_k^{AD}$ and bootstrap $\hat{\gamma}_k^{*CL}$ values of the predictors

$\hat{\alpha}_i^{LD}$	$\hat{\alpha}_i^{CC}$	$\hat{\alpha}_i^{*LD}$	$\hat{\alpha}_i^{*CC}$
1483	1456	1483	1453
1431	1439	1431	1436
1558	1542	1557	1538
1951	1884	1944	1880
2149	2055	2135	2051
2148	1713	2121	1709

Table 7 The original (left) and bootstrap (right) values of the expected ultimate loss

5 Conclusion

When the stochastic version of the Bornhuetter - Fergusson principle is used for calculation of the claims, the problems caused by lack of knowledge of the probability distribution of the data arise. The origin of mentioned problems is usually the small number of the data. In such cases the bootstrap method can be applied, principle of which is relatively simple without presumption of knowledge of probability distributions. The authors presented three versions of application of Bornhuetter - Fergusson principle for claims reserves calculation with the concrete data in this article. The results shown in the tables 3 - 7 don´t differ significantly each other; we can conclude that the application of the bootstrap for prediction of the claim reserves provides reasonable results and thus it can be used in practice. Of course, as with all the methods, it is necessary to confront the results with other methods.

Acknowledgements

The article is result of the project GAČR 402/09/1866 Modelling, simulations and management of insurance risk.

References

[1] Bornhuetter, R. L., and Ferguson, R.E.: The Actuary and IBNR. In: *Proceedings of the Casualty Actuarial Society 59*, 1972, 181-195.
[2] Fecenko, J.: *Neživotné poistenie.* Ekonóm, Bratislava, 2006.
[3] Linda, B., Kubanová, J., and Jindrová, P.: Insurance reserves estimation by bootstrap. *Scientific Papers of the University of Pardubice* **21** (2011), 127-138.
[4] Mack, T.: Parameter Estimation for Bornhuetter-Ferguson. *Casualty Actuarial society Forum,* Fall 2006, 141-157
[5] Pacáková, V.: Modelling and Simulation in Non-life Insurance. In: *Proceedings of the 5th International Conference on Applied Mathematics, Simulation and Modelling* (Mastorakis, N. at al., eds.). WSEAS Press, Corfu Island, Greece, 2011, 129 - 133.
[6] Pacáková, V.: *Aplikovaná poistná štatistika.* Iura Edition, Bratislava, 2004.
[7] Stehlík, M., Potocký, R., Waldl, H., and Fabián, Z.: On the favorable estimation for fitting heavy tailed data. *Computational Statistics* **25**, (2010), 485–503.
[8] Schmidt, K. D.: Methods and Models of Loss Reserving Based on Run-Off Triangles: A Unifying Survey. *Casualty Actuarial society Forum,* Fall 2006, 269-317
[9] Schmidt, K. D., and Zocher, M.: The Bornhuetter-Ferguson principle. *Variance* 2 (2008), 85-110.
[10] Wüthrich, M.V., and Merz, M.: *Stochastic claim reserving methods in insurance*. John Wiley & Sons Inc., Chichester, 2008.

Contribution to financial distress and default modeling and new 2-D aggregated model – SME case studies

Ladislav Lukáš [1]

Abstract. Financial distress and default modeling represents large and important field of economic research topics. This paper concerns with general structure of such models in form of linear or affine forms built upon data available from standard accounting reports. Such general approach helps both to elucidate connections among well-known models, e.g. Altman, Ohlson, Neumaier and other ones, as well as building new models. Using logit transformation we get probabilistic version of Altman Z-score which thus may form one component of a new 2-D aggregated model, whereas the second component is constituted by Ohlson O-score in proper form. These models are used both for extensive analysis and short-term prediction of firm's state. All models are implemented by Mathematica modules and important algorithmic details are presented. Finally, we present and discuss some illustrative results of SME case studies, as well.

Keywords: Default models, financial distress, firm behaviour, firm performance, prediction of firm's state.

JEL Classification: C33, C81, D21, D81, L25
AMS Classification: 91B82, 91G50

1 Introduction

Financial distress and default modeling represents large and important field of economic research topics. This paper concerns with general structure of such models in form of linear or affine forms built upon data available from standard accounting reports. Such general approach helps both to elucidate connections among well-known models, e.g. Altman, Ohlson, Neumaier and other ones, as well as building new models. There is well known that financial statements are formulated under the going-concern principle, which assumes that firms would not go bankrupt in general, however that could be caused also by new business circumstances, market environment and/or any another economic related impacts.

Our analysis of accounting-based default indicators is based on the traditional Altman's Z-score, Ohlson's O-score and Neumaier index of trustworthiness IN05. All such indicators are calculated using fiscal year-end data, and computed this way they do not represent bankruptcy probabilities. However, they can be turned into probabilities using the logistic transformation.

2 Structure of accounting-based default models

At present, there are two big groups of default models available. The first and older one is based upon accounting reports. The newer one is generally based upon company market pricing, and the indicators are constructed using methods of financial engineering, e.g. Black-Scholes option pricing method. We shall concern ourselves with the first group, and we assume that we have at our disposal time series data form standard reports, i.e. balance sheets, profit-and-lost reports and cash-flow reports.

Let us assume we have k different fiscal data available per year from these standard reports. Hence, we are able to build $C(k,2)=k(k-1)/2$ couples thereof, which may form ratios, i.e. dimensionless quantities. There is evident that not all of them are popular and used in default models. So, having a couple (a,b) we may set either ratio a/b or b/a as well, which can serve the same purpose since they are turned each other by reciprocal transformation, e.g. $b/a = (a/b)^{-1}$.

Let denote $n=C(k,2)=k(k-1)/2$, hence we may interpret all available ratios from accounting reports as points in the space R^n. Such construction enable us to understand particular default models in a unique way as small dimensional sub-spaces R^m, i.e. hyperplanes in R^n, where $m < n$. Hence, we may write a general form of such models using scalar product formula (1)

$$(\mathbf{c}, \mathbf{x}) , \quad \mathbf{c} = (c_1, c_2, ..., c_m)^{\mathrm{T}}, \ \mathbf{x} = (x_1, x_2, ..., x_m)^{\mathrm{T}} \tag{1}$$

[1] University of West Bohemia, Faculty of Economics, Husova 11, 30614 Plzeň, lukasl@kem.zcu.cz .

where $\mathbf{c}$ is m-dimensional weight vector given, and $\mathbf{x} \in R^m$, $R^m \subset R^n$, and ^{T} means vector transposition.

Since particular models use different financial ratios, which means they are defined in different subspaces R^m and use different weight vectors, we can identify and code them in an unique way as follows – Altman's Z-score $({}_A\mathbf{c},{}_A\mathbf{x})$, Ohlson's O-score $({}_O\mathbf{c},{}_O\mathbf{x})$, and Neumaeirs' index of trustworthiness IN05 by $({}_N\mathbf{c},{}_N\mathbf{x})$, while identifying the corresponding subspaces to be ${}_sR^m$, with s = A,O,N.

There is a general way how to implement such default models – first, to calculate the value of the indicator, say $\gamma(\mathbf{x}) = (\mathbf{c},\mathbf{x})$, and then to check whether the computed value belongs or not to so-called shadow area, denoted D in general, in order to emphasize its role do mark an inclination to default and to derive a statement of company's health. Another point of view says that $\gamma(\mathbf{x})$ can be tackled as a linear mapping $\gamma : R^m \to R$, too. Since the discussed models are well known we give just their definitions.

Altman's model (version 1983): $Z = \gamma({}_A\mathbf{x}) = ({}_A\mathbf{c},{}_A\mathbf{x})$, where ${}_A\mathbf{c} = (0.717, 0.847, 3.107, 0.420, 0.998)^T$,

${}_A\mathbf{x} = ({}_Ax_1, {}_Ax_2, ..., {}_Ax_5)^T \in {}_AR^5$, with ${}_Ax_1$ = WC/TA, ${}_Ax_2$ = EAR/TA, ${}_Ax_3$ = EBIT/TA, ${}_Ax_4$ = VE/TL, ${}_Ax_5$ = S/TA,

WC..working capital, TA..total assets, EAR..retained earnings, VE..market value of equity, S..sales, TL..total liabilities,

${}_AD$ = [1.20, 2.90] .. Altman's (version 1983) shadow zone.

Altman's model (version 1995): ${}_{95}Z = \gamma({}_{A95}\mathbf{x}) = ({}_{A95}\mathbf{c},{}_{A95}\mathbf{x})$, where ${}_{A95}\mathbf{c} = (6.56, 3.26, 6.72, 1.05)^T$,

${}_{A95}\mathbf{x} = ({}_{A95}x_1, {}_{A95}x_2, ..., {}_{A95}x_4)^T \in {}_{A95}R^4$, with ${}_{A95}x_1 = {}_Ax_1$, ${}_{A95}x_2 = {}_Ax_2$, ${}_{A95}x_3 = {}_Ax_3$, ${}_{A95}x_4 = {}_Ax_4$,

${}_{A95}D$ = [1.20, 2.60] .. Altman's (version 1995) shadow zone.

Ohlson's model: $O = \gamma({}_O\mathbf{x}) = {}_Oc_0 + ({}_O\mathbf{c}, {}_O\mathbf{x})$, where ${}_Oc_0$ = -1.32 is an additive constant,

${}_O\mathbf{c} = (6.03, -1.43, 0.08, -2.37, -1.83, 0.285, -1.72, -0.52)^T$,

${}_O\mathbf{x} = ({}_Ox_1, {}_Ox_2, ..., {}_Ox_9)^T \in {}_OR^9$, with ${}_Ox_1$ = ln(TA/GDP_PI), ${}_Ox_2$ = TL/TA, ${}_Ox_3$ = WC/TA, ${}_Ox_4$ = CL/CA,

${}_Ox_5$ = NI/TA, ${}_Ox_6$ = FFO/TL, ${}_Ox_7$ = if (EAR in last 2 years < 0) then 1 else 0,

${}_Ox_8$ = if (TL > TA) then 1 else 0, ${}_Ox_9 = (EAR_t - EAR_{t-1})/|EAR_t - EAR_{t-1}|$,

GDP_PI..GDP price level index, CL..current liabilities, CA..current assets, NI..net income.

Neumaiers' model (version 2005): $N = \gamma({}_N\mathbf{x}) = ({}_N\mathbf{c}, {}_N\mathbf{x})$, where ${}_N\mathbf{c} = (0.13, 0.04, 3.97, 0.21, 0.09)^T$,

${}_N\mathbf{x} = ({}_Ax_1, {}_Ax_2, ..., {}_Ax_5)^T \in {}_NR^5$, with ${}_Nx_1$ = TA/FC, ${}_Nx_2$ = EBIT/CI, ${}_Nx_3$ = EBIT/TA, ${}_Nx_4$ = S/TL, ${}_Nx_5$ = CA/CL,

FC..foreign capital, CI..cost interest, ${}_ND$ = [0.90, 1.60] .. Neumaiers' shadow zone.

Different shadow zones which belong to particular default models provide an evident possibility to map each one to other one with linear mappings. First, we define mapping M which maps ${}_ND \to {}_AD$ by (2)

$$M: \xi \to \eta = \alpha_0 + \alpha_1\xi, \quad \xi \in {}_ND, \ \eta \in {}_AD, \tag{2}$$

where mapping constants α_0, α_1 are determined from corresponding interpolation conditions. An inverse mapping M^{-1} realizing ${}_AD \to {}_ND$ is given by (3). The mapping constants β_0, β_1 are determined from the same interpolation conditions defined in inverse setting.

$$M^{-1}: \eta \to \xi = \beta_0 + \beta_1\eta, \quad \xi \in {}_ND, \ \eta \in {}_AD. \tag{3}$$

3 Numerical results – case studies

All case studies analyzed in this paper were performed by Mathematica notebook. Our present experience shows that sw system Mathematica, Wolfram Research Inc., is ideally suited for such tasks since it enables both numerical calculations to be accompanied by extensive creation of graphs and their output and/or export. We have selected various companies ranging from larger one, e.g. Czech Airlines, till relatively small ones ranging into SME framework.

There is always recommendable to check the input time series data by graphical outputs in order to inspect any errors, first. Though we have done them in all our case studies, they are not given here because of limited paper space. We refer to [7] for corresponding source data and more details relating companies being analyzed.

Case study 1 – ČSA (Czech Airlines). We have excerpted data from public available standard accounting reports for ten years long period, from 2001 till 2010. First, we have run Altman's model using the version from 1983. Further we amend the analysis running Z-score, version 1995, too. The corresponding results are given on Figure 1 and 2. In both cases the shadow zones are depicted. In particular on Figure 2, there is evident that the Z-score model ver.1995 gives sharper results as to the financial distress and prospective default of company. The very critical year was 2009. However, the next year 2010 shows extreme recovery. We have completed such analysis with Ohlson's model. The corresponding result is given on Figure 3.

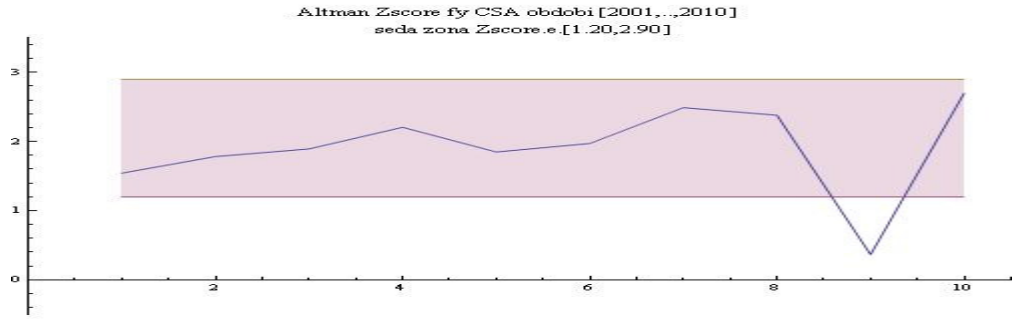


Figure 1 ČSA – Z-score version 1983, period 2001-2010

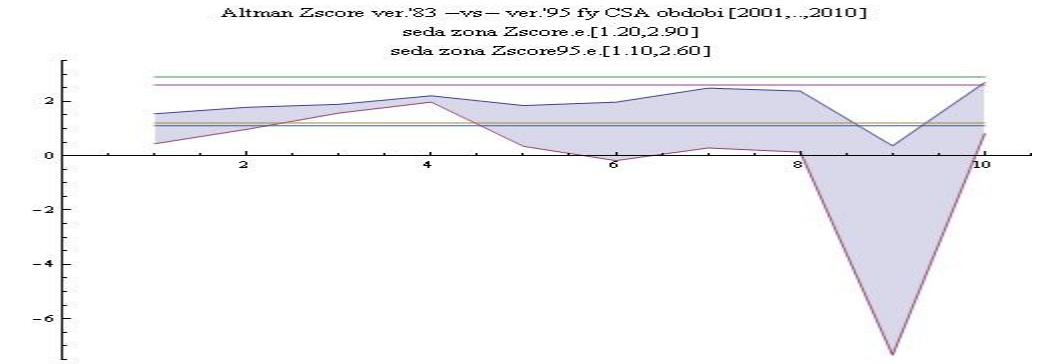


Figure 2 ČSA – Z-score versions 1983 and 1995, period 2001-2010

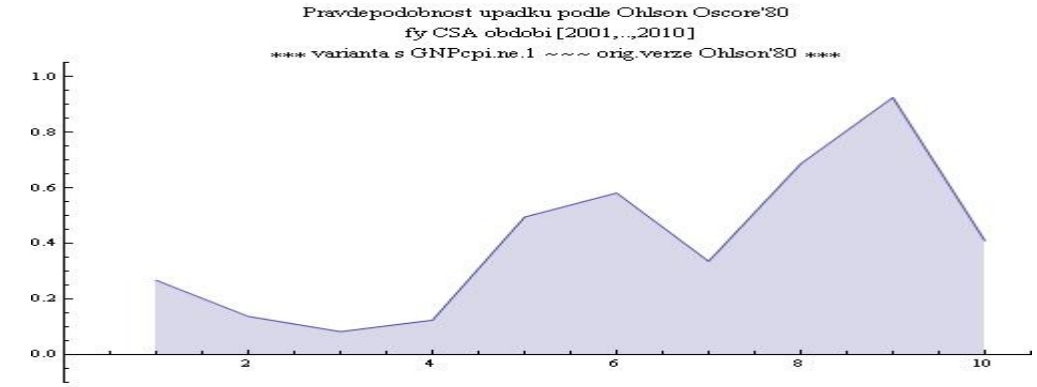


Figure 3 ČSA – Probability of default by Ohlson's model, period 2001-2010

Case study 2 – Dajbych s.r.o (Ltd., Off-road cars dealer). The company fits into SME range. We have collected data using same procedure as in the previous case for twelve years long period, from 2000 till 2011. Since there is a typical Czech SME company we have used both Altman's model and the Neumaeirs' model IN05 as well, in order to inspect the prospective differences therein. First, we have run both models individually, but further we adopted linear mappings M and M^{-1} given by (2) and (3), in order to investigate their applicability. The results are given on Figure 4, 5 and 6.

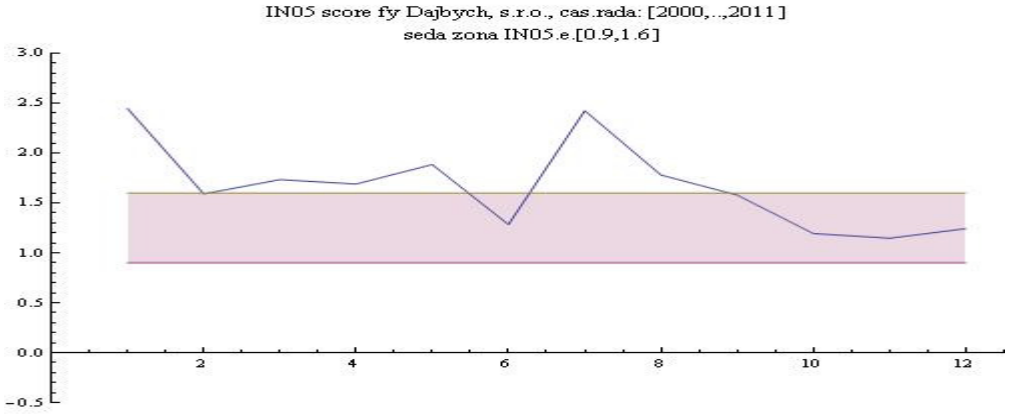


Figure 4 Dajbych s.r.o. – IN05 index, period 2000-2011

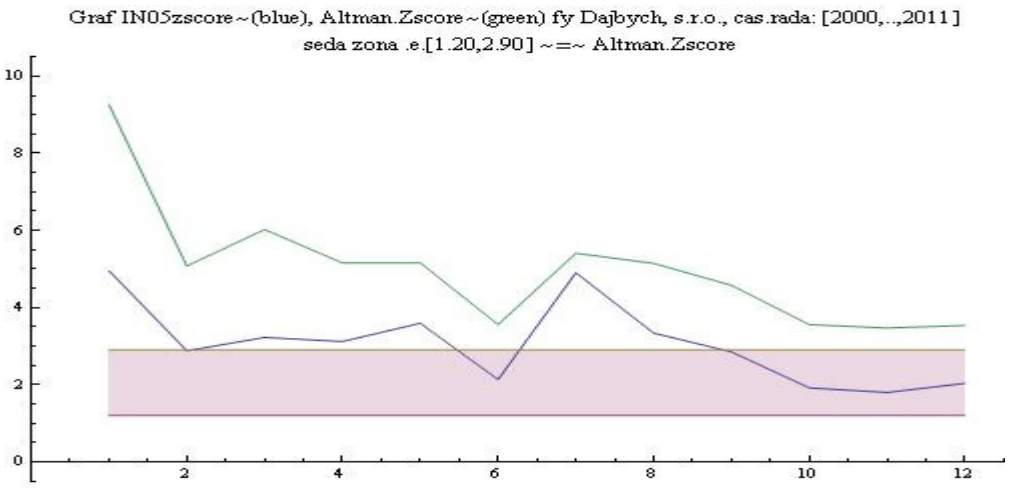


Figure 5 Dajbych s.r.o. – IN05 period 2000-2011, mapping M: ${}_{N}D \rightarrow {}_{A}D$

Figure 6 shows also two graphs as Figure 5 does. The first one gives values of IN05 for the period 2000-2011 which can be identified by comparing with the graph on Figure 4. However, the second graph shows values of Altman's Z-score ver.1983 mapped onto Neumaeirs' scale. The procedure yields the following values:

{3.41004,1.68502,2.07444,1.71769,1.71866,1.05981,1.82129,1.71313,1.47823,1.05894,1.0229,1.05042}.

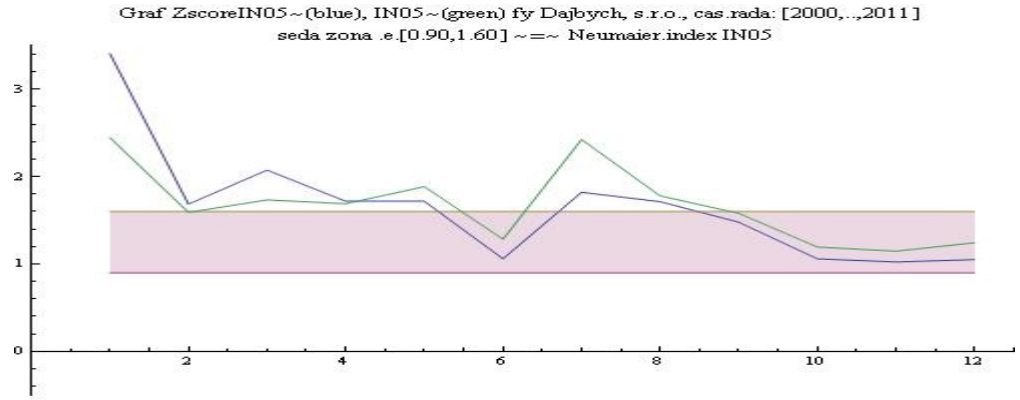


Figure 6 Dajbych s.r.o. – IN05 period 2000-2011, mapping M^{-1}: ${}_{A}D \rightarrow {}_{N}D$

Comparing graphs on Figure 5 and 6 gives interesting results related company's financial distress and default warning. Figure 5 shows that Z-score does not detect any distress during whole period. On the contrary, the mapped IN05 plunges into ${}_{A}D$ at least partially. Applying inverse mapping M^{-1} we can conclude the both indicators issue warning signals for company health danger. The last three years, i.e. 2009, 2010 and 2011, were rather difficult for the company to survive. Hence, prediction is to be calculated urgently. We have used three prediction models, two linear ones and the third quadratic one, both given by (4), where corresponding coefficients are

$$f_1(\xi) = a_0 + a_1\xi\,, \quad f_2(\xi) = a_0 + a_1\xi + a_2\xi^2. \tag{4}$$

determined by usual interpolation conditions. Prediction results are depicted on Figure 7.

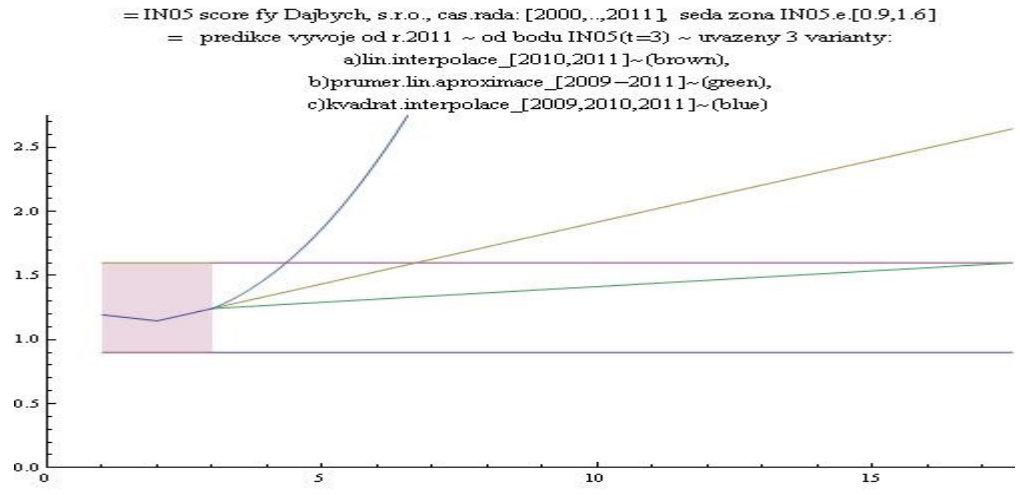


Figure 7 Dajbych s.r.o. – IN05 predictions by (4) built from period 2009-2011

Finally, we propose new aggregated 2-D default model with the range [0,1]x[0,1], which components (p_1,p_2) are formed by probability measures built from A-score and O-score by logistic transformations (5), respectively.

$$p_1(\xi) = 1/(1+\exp(-\xi)),\ \xi = b_1 + b_2\,\gamma({}_{A}\mathbf{x})\,, \quad p_2(\eta) = 1/(1+\exp(-\eta)),\ \eta = \gamma({}_{O}\mathbf{x}) \tag{5}$$

The coefficients b_1, b_2 are determined by interpolation conditions relating ${}_{A}D$ with selected quantile period, e.g. [0.05, 0.95]. Hence, using expressions (5) we may define mapping F: $(\gamma({}_{A}\mathbf{x}), \gamma({}_{O}\mathbf{x})) \rightarrow (p_1,p_2) \in [0,1]\text{x}[0,1]$. The very first result is given on Figure 8 with four iso-quantile lines.

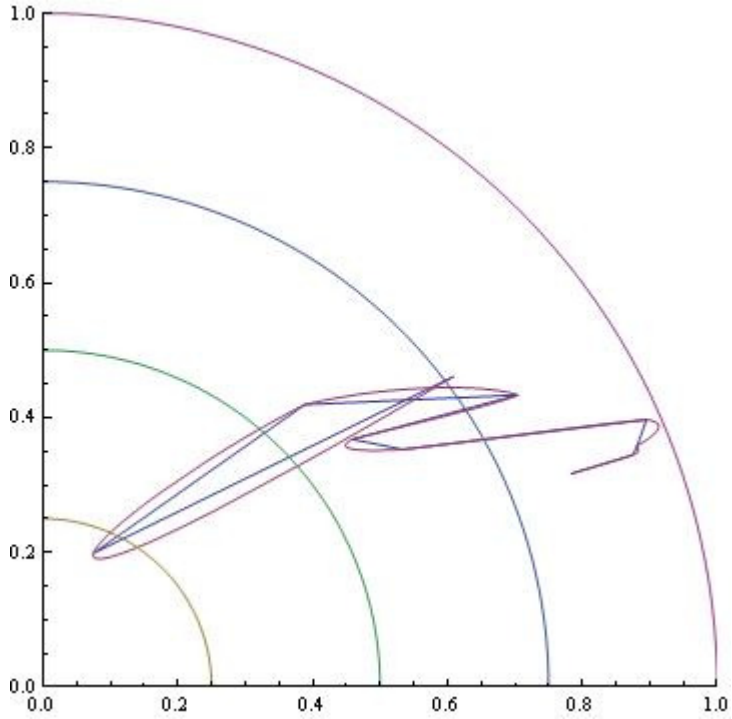


Figure 8 ČSA – trajectory of survival probabilities $(1-p_1, 1-p_2)$, period 2001-2010

4 Conclusion

The paper presents general approach for construction default models based on fiscal data reported in accounting statements available in public, and discusses the well-known models within that framework, too. Two interesting topics have been open. First, mapping of shadow zones of different models each other is presented in detail. Second, the new aggregated 2-D model based on probabilistic measures is discussed, too. Further research will be focused both on empirical verification of the new model and advanced theoretical development, too.

References

[1] Altman, E.: Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance* **23** (1968), 589–609.
[2] Fama, E. and French, K.: Industry Costs of Equity. *Journal of Financial Economics* **43** (1997), 153–193.
[3] Kubíčková, D., and Kotěšovcová, J.: *Financial Analysis* (*Finanční analýza*, in Czech). Prague, VŠFS – EUPRESS, 2006, 125 pp, ISBN 80-86754-57-X.
[4] Muchová, D.: *Analysis of company during economic crisis* (*Analýza vývoje podniku v době hospodářské krize*, in Czech). Thesis, Univ. of West Bohemia in Pilsen, ZČU/FEK, Pilsen, 2012.
[5] Ohlson, J.: Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research* **19** (1980), 109–131.
[6] Sedláček, J.: *Corporate Financial Analysis* (*Finanční analýza podniku*, in Czech). Brno, Computer Press, 2009, 154 pp., ISBN 978-80-251-1860-6.
[7] Tolarová, I: *Identification of company crisis development* (*Identifikace krizového vývoje podniku*, in Czech). Thesis, Univ. of West Bohemia in Pilsen, ZČU/FEK, Pilsen, 2012.
[8] Zmijewski, M.: Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, Supplement, 1984, 59-86.
[9] http://eprints.covenentuniversity.edu.ng/73/1/A_ Comparative_Analysis_Of_The_Effectiveness_Of_Three_ Solvency_Management_Models.pdf.
[10] http://www.tamut.edu/jolb/Scholar/2010Summer/2010Wang-Campbell.pdf.
[11] http.//www.indianresearchjournals.com/pdf/IJMFSMR/2012/April/10.pdf.
[12] http://neumann.hec.ca/gestiondesrisques/Simard-projet-sop.pdf.
[13] http://arxiv.org/ftp/arxiv/papers/1103/1103.2110.pdf.

Modelling Population Dynamics for Archaeological Simulations

Tomáš Machálek[1], Kamila Olševičová[2], Richard Cimler[3]

Abstract. We intend to apply the agent-based modelling and social simulation to study the complexity of the Celtic society in Central Europe in the late Iron Age. Key aspects which form the complexity of the society are settlement forms, demography, the scale of work specialization etc. Our objective is to determine under what conditions the collapse of the Celtic society might have taken place. The paper presents particular results, especially the explanatory population dynamics model which was implemented in NetLogo. The model is based on specific domain knowledge and general demographic suppositions about birth rates, mortality and immigration. The model allows archeologists to simulate the time series of available workforce and actual consumption of the population living in the settlement agglomeration Staré Hradisko. The population dynamics model is essential for further husbandry management model. The hypotheses related to extensive and/or intensive agricultural practices should be verified.

Keywords: agent-based model, archaeology, NetLogo, population dynamics, social simulation

JEL Classification: J10
AMS Classification: 68U20

1 Introduction

The *agent-based modelling* and *social simulation* is being applied in many domains including the archaeology for last ten years successfully. The agent-based modelling is characterized by the use of microspecifications (agents, rules, environments) that are sufficient to generate the macrostructures of interests. The creation of social simulation consists of (1) *conceptual modelling* of processes and objects of real world according to research questions and hypotheses, (2) *implementation* of models, (3) running experiments with *data analysis*, (4) *verification* and optional revisions of the model, (5) *sharing results* and (6) *reproducing the simulation*.

Benefits of this *generative social science* approach have been highlighted in [4]. Typically there are two types of research results: (1) interdisciplinary domain models and (2) simulation platforms, methodologies and guidelines. For example Altaweel's realistic models of ancient Mesopotamian civilization are presented simultaneously with the ENKIMDU chassis [1]. The simulation of behavioral patterns of early hominids [5] is described together with the demonstration of applying the Overview-Design-Details protocol [6]. The well-known investigation of cultural collapse of ancient Anasazzi civilization [2] is discussed along with the replication the model in NetLogo [7].

We intend to apply agent-based models and social simulations to analyze the conditions and circumstances of the development and collapse of Celtic society in Central Europe in the late Iron Age. Key aspects which form the complexity of the society are: settlement forms (oppida), demography, agricultural practices, producers/consumers ratio in society and the scale of work specialization, local and distant interactions (trade and exchange, monetary economy) and others [3].

The paper presents particular results related to the explanatory population dynamics modelling. The model is based on domain knowledge provided by experts (archaeological excavations of various oppida, regional landscape studies at Staré Hradisko, demographical studies and assumptions, life-expectancy tables [9] etc.). The population dynamics model is essential for further investigation of the *carrying capacity* of settlement agglomerations.

[1] University Hradec Králové, Faculty of Informatics and Management, Rokitanského 62, Hradec Králové 50003, tomas.machalek@uhk.cz
[2] University Hradec Králové, Faculty of Informatics and Management, Rokitanského 62, Hradec Králové 50003, kamila.olsevicova@uhk.cz
[3] University Hradec Králové, Faculty of Informatics and Management, Rokitanského 62, Hradec Králové 50003, richard.cimler@uhk.cz

The organization of the rest of the paper is as follows. The overview of our population dynamics model is provided in chapter 2, the NetLogo implementation and interfaces are presented in chapter 3, further research directions are discussed in conclusion.

2 Overview of the Model

A population dynamics model answers the question how a population is going to change given (1) its current status, i.e. the total number of individuals and the composition of the population in terms of old/young individuals) and (2) the environmental conditions that the population is exposed to. The population change is characterized by births, deaths, immigration and emigration. It is assumed that all populations grow (or decline) exponentially (or logarithmically) unless affected by other forces [8].

Following basic assumptions and requirements related to our model were provided by archaeologists:

- The period of 100-120 year has to be simulated.
- The initial number of inhabitants in the settlement is between 500 and 800. The maximum number at the end of simulation is between 2000-5000.
- There are three types of peasant families: large size (aprox. 20 members), medium size (aprox. 10 members) and small size (4-6 members).
- A peasant family has got 2 adults, 1-3 children and 1 elder. Correspondingly the medium size family has got 4 adults and the large family has got 8 adults.
- In other words, the large family consists of approximately 7 infants (1 suckling, 3 todlers, 3 up to 10 years), 3 older children (10 - 14 years), 2 young adults (15 - 19 years), 5 adults and 3 elderly.
- Social roles (basic family, nobility, servants, slaves etc.) are ignored.
- Personal histories of individuals (marriages, children, siblings etc.) are ignored.
- The fertile age of women is 15-49 years and the fertility rate is 5.1. More than two children rarely survived infancy.
- The probabilities to die are defined in five-year life tables [9], complete one-year life tables could be estimated using demographic methods.
- Mild annual growth of population in settled conditions and with respect to high child mortality is estimated to 0.2%.
- Very probable immigration should be considered; typically one small or medium size household is established every five years.
- Emigration should not be taken into account at the moment. Massive emigration is one of possible causes of disappearance of settlement population and it will be investigated separately.
- The workforce is expressed as the number of men between 15-49 years who are able to plough.
- The daily calories requirements are defined (1360 for toddlers, 2000 for small children and elderly, 2500 for boys between 10 and 14 etc.).

The model inputs are limited to:

- Slider for setting the initial *number-of- years-to-be-simulated* (between 100-120),
- Slider for setting the initial *number-of-inhabitants* (between 500 and 800),
- Slider for setting the initial *number-of-large-families* (between 20 and 40),
- Selection list of available *life-expectancy-table*,
- Slider to setting *the birth-probability*,
- Slider for setting the *immigration-probability*.

There are two output variables:

- time series of *actual-consumption* of population (calories),
- time series of *actual-workforce* (number of men between 15-49).

3 Implementation and interfaces

The model is implemented in NetLogo, a multi-agent programmable modeling environment [10]. Two types of agents are defined:

- *household-agent* – has got its *type* (large, medium, small),
- *inhabitant-agent* – has got its *age* and *gender*, is linked to the *household-agent.*

Auxiliary variables were added for monitoring characteristics of each *household-agent* and of the whole population: *num-of-sucklings, num-of-todlers, num-of-children, num-of-older-children, num-of-young-adults, num-of-adults, num-of-elder*, and summarizing *num-of-inhabitants*, *actual-workforce* and *actual-consumption* inform us about the structure the population.

The initialization of the model involves creation of appropriate *household-agents*. The number of small and medium households is derived from the initial *number-of-inhabitants* and the initial *number-of-large-families*. For each *household-agent* the set of *inhabitant-agents* is created. Random numbers from the normal distribution with appropriate mean and standard deviation are generated to obtain particular numbers of sucklings, toddlers, children etc. for each household.

The current population structure of households is visualized (Fig. 1). Each circle diagram represents one household, i.e. the total number of family members (in the center of the diagram) and the ages of individuals ascending clockwise. The diagrams are updated periodically as well as the graph of the population growth (Fig. 2).

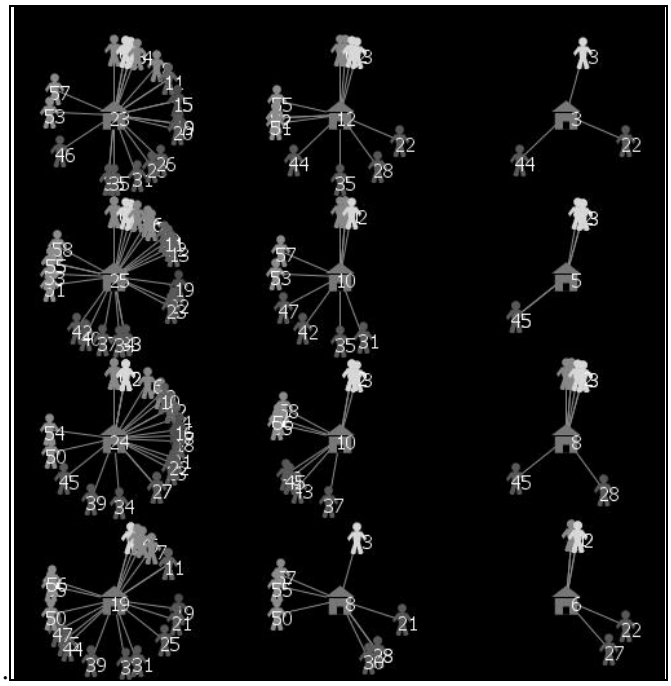


Figure 1: Visualization of households and inhabitants (part of the screenshot)

The model is comprised of nested cycles characterized by updating the state of *household-agents* and *inhabitants-agents*. For each year of the simulation, three procedures repeat:

- *birth rate procedure* – the probability formula for adding a newborn *inhabitant-agent* is based on current number of fertile women in the household and the *birth-probability* parameter,
- *mortality rate procedure* – the probability formula for removing *inhabitant-agent* from the model is based on *life-expectancy tables*,
- *immigration rate procedure* – the probability formula for adding a new small or medium size family is based on the *immigration-probability* parameter.

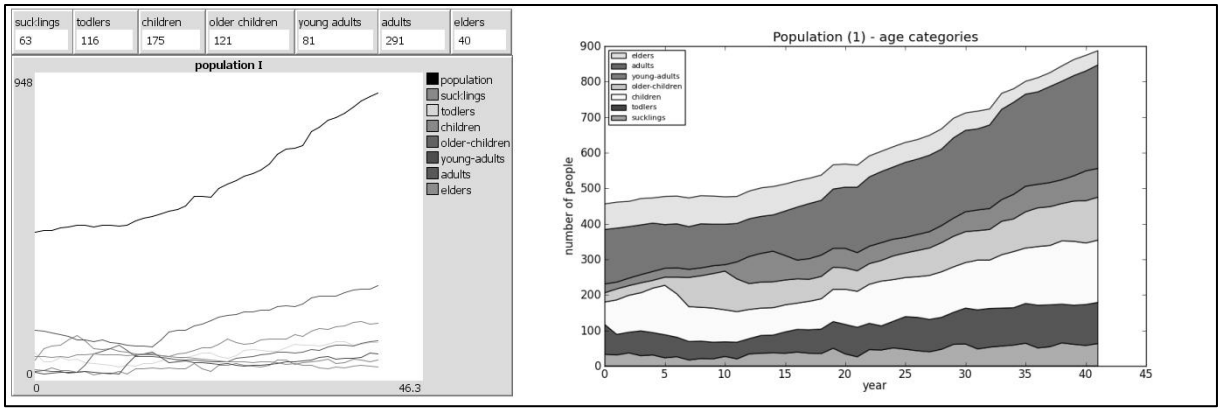


Figure 2: Population structure

4 Conclusion

Our agent-based population dynamics model of the Celtic oppidum enables experimenting with birth rates, mortality and immigration rates. The formulas can be further refined according to expertise (e.g. different calorie requirements tables or interpolations of life-expectancy tables can be inserted, parameters of normal distributions in setup procedure can be particularized etc.). If necessary the life-cycle of the *household-agent* could be enhanced by defining social roles and personal histories of *inhabitant-agents* and other features that would make the model more realistic (but also highly complex).

The outputs of the population simulations are intended to become inputs of the agricultural model. The time series of *available-workforce* and *actual-consumption* can help us to estimate the *carrying capacity* of oppida, (the environment's maximal load), one of explanatory variables important for understanding the circumstances of the collapse of Celtic society. It is known that Celts received 80 percent calories from cereals. It means that the effective husbandry practices on more or less fertile arable land in walking distance around the settlement were crucial. The archeologists assume some combination of extensive and intensive husbandry and the agricultural model should help them to verify particular hypotheses about the Staré Hradisko settlement area.

The agent-based models applied in social sciences are sometimes criticized for oversimplifications and the lack of flexibility. We hope that with two relatively independent models it is possible to avoid this risk. The population dynamics model operates with *individual-agents* and *household-agents* while the agricultural model should conceptualize the land use, i.e. its basic element is a *land-patch-agent*.

Further research will be focused on:

- refinement of the population dynamics model,
- development of the agricultural model,
- development of scripts for better presentation and visualization of data exported from NetLogo.

Acknowledgements

The research described in the paper has been funded by the Czech Science Foundation (grant no P405/12/0926 *Social modelling as a tool for understanding Celtic society and cultural changes at the end of the Iron Age*) and by FIM UHK (students' research grant).

References

[1] Altaweel, M. Investigating agricultural sustainability and strategies in northern Mesopotamia: results produced using socio-ecological modeling approach. *Journal of Archaeological Science*, **35** (2008), 821-835.
[2] Axtell, R.L. et. al.: Population Growth and Collapse in a Multi-Agent Model of the Kayenta Anasazi in Long House Valley. *Proceedings of the National Academy of Sciences*, **99** (2002), 7275-7279.
[3] Danielisová, A.: *The oppidum of České Lhotice and its hinterland*. Ph.D. dissertation, unpublished, 2008.
[4] Epstein, J.M., Axtell, R.: *Growing Artificial Societies: Social Science from the Bottom Up*. MIT Press, 1996.
[5] Griffith, C.S., Long, B.L., Sept, J.M.: HOMINIDS: An agent-based spatial simulation model to evaluate behavioral patterns of early Pleistocene hominids. *Ecological Modelling*, **221** (2010), 738-760.
[6] Grimm, V. et. al.: The ODD protocol: A review and first update. *Ecological Modelling*, **221** (2010), 2760-2768.
[7] Janssen, M.: Understanding Artificial Anasazi. *Journal of Artificial Societies and Social Simulation*, **12** (2009), (4) 13.
[8] de Roos, A.M.: *Modeling Population Dynamics*. Course material, http://staff.science.uva.nl/~aroos/downloads/pdf_readers/syllabus.pdf. Universiteit van Amsterdam, 2011.
[9] Saller, R.: *Patriarchy, Property and Death in the Roman Family*. Cambridge, 1994.
[10] Wilensky, U.: *NetLogo*. http://ccl.northwestern.edu/netlogo/ Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL, 1999.

Managerial D-M: Measuring of risk scenes and tools of their reducing

Dusan Marcek[1], Alexandra Kotillová[2], Michal Ulbricht[2]

Abstract. The paper is concerned with measuring and assessment of risk scenes in managerial decision-making (D-M). It builds upon the uncertainty of economic information, which is converted into the concept of risk scene expressed in terms of probability and using confidence intervals of the predicted quantities. The paper explains the relation of a degree of risk expressed by the classical information measure, bit, by the concept of confidence intervals, or possibly by the standard deviation. The risk connected with managerial decision-making is modeled using probability models and understood as a statistical term of the expected value between two extreme states of decision. Forecasting systems are applied which are based on the latest statistical theory and artificial neural networks. The degree of risk is assessed. The impact of these methods to risk reduction is judged in managerial decision-making.

Keywords: confidence intervals, uncertainty, entropy, forecasting models, neural networks, managerial D-M, risk scene assessment.

JEL Classification: C13, C45, D81
AMS Classification: 90B50

1 Introduction

An important sphere of information necessary for management of production processes on all managerial levels is the information about the future development of quantities expressed quantitatively, which is used to characterize the state and the development of the object or process. Evidence shows that it is possible to make this information more precise by a suitable choice and use of forecasting models based on statistical methods, soft computing and artificial intelligence methods. In comparison with the manager's expert estimates, these models based on statistical and soft computing methods or artificial intelligence methods are capable of providing information in the form of forecasts of quantities with an acceptable degree of uncertainty. The manager using these forecasts is able to make better decisions, i.e. such decisions whose risks in achieving targets are minimized.

Most of the real systems can only be described incompletely, i.e. with information which cannot be formally expressed by unequivocally set parameters. This is uncertain information then. In practice, there are mainly two types of such information. According to the first type, uncertain information makes it impossible to exactly determine the future behavior of the examined system. This type of uncertainty is called stochastic, and it can usually be modeled using the probability theory. The second type of uncertainty is connected with the description or formulation of the actual meaning of the phenomena or statements about them. This is semantic uncertainty. Natural language words semantics with uncertainty, i.e. with meanings of words and individual statements not being exact, is typical of natural language. This uncertainty has the character of possibility rather than probability, and it is most often modeled by fuzzy systems.

One of the approaches to understanding uncertainty in forecasting models is understanding it as the standard deviation σ of the forecasted quantity or process [7]. The standard deviation as a degree of uncertainty, or risk, of forecasted quantity values estimates is equivalent to the statistical degree of accuracy of the forecast defined as Root Mean Square Error of the forecast. It need to be stated that the standard deviation does not reflect entropy in its true substance as uncertainty which is indicated in bits (binary digits). On the other hand, uncertainty is closely related to how precise are the estimates of the future values of quantities that managers have at their disposal. This view of uncertainty does not articulate it in its true sense, however, it expresses very well its inner essence and the mutual relation of entropy and D-M.

The issue of measuring risk in management and its accompanying phenomena is divided into four chapters in the present paper. Chapter two is devoted to characterizing risk and its manifestation in decision-making in uncertainty conditions. In the third chapter, a diagram of an uncertainty reduction procedure in the manager's decision-making is designed and characterized. In the fourth chapter, risk reduction with the use of forecasting

[1] Research Institute of the IT4Innovations Centre of Excellence, Silesian University in Opava, Bezruc. Square 13, 746 01 Opava, Czech Republic, email: Dusan.Marcek@fpf.slu.cz
[2] University of Žilina, Faculty of Management Science and Informatics, Department of Informatics, Univerzitná 8215/1, 01026 Žilina, Slovakia, email: kotillova@gmail.com, michal.ulbricht@fri.uniza.sk

models based on the classical (statistical) methods and models based on artificial intelligence is documented and assessed. Chapter five summarizes the main topics and results.

2 The relation between D-M with uncertainty and D-M with risk

As it was mentioned in the preceding chapter, within the managerial decision-making process uncertainty indicates the degree of risk of achieving targets. Decision-making on the level of lower management usually involves theories and tools such as linear and non-linear programming, dynamic programming, game theory, queuing theory, inventory theory, probability theory, renewal theory, graph theory etc. Decision making on the level of top management is significantly influenced by time. Top management uses tools not only from management but also from other science branches such as mathematical statistics, fuzzy set theory, econometrics, operational research, etc. Top managers use these tools to obtain the most precise estimates of the future development of quantities and processes possible. These estimates represent important information on which managers base their decisions. Specific choice of tools and models for decision-making depends on whether the manager has precise and complete or imprecise and incomplete information at their disposal. The complexity of managerial decision-making relates to decision-making with incomplete information.

Stochastic uncertainty is concerned with the category of the probability risk, which is determined as a scene in the future associated with the specific adverse incident that we are able to predict it using probability theory and a lot of data [3]. In this manuscript, we will concern with this type models, which may be described as follows. Let D be a managerial prediction system including explanatory variables V to explain the behavior of the variable to be forecast, and faults represented as forecast errors e_t in time $t = 1, 2,\ldots n$. A risk function R in term of the conceptual model D for having a risk scene can be represented as

$$R = D(V,\ e_t), \qquad t = 1,\ 2,...n\ .$$

To assess the managerial prediction risk R we apply different forecasting models which parameters are estimated by statistical tools.

As far as decision-making with risk is concerned, this is the case of decision-making where actual information about real systems is uncertain, and it is not important if the uncertainty is caused by incomplete information about the system's behavior, or if it is semantic uncertainty. In the further text, in accordance with, the risk connected with managerial decision-making will be modeled using probability models and understood as a statistical term of the expected value between two extreme states of decision, i.e. with full uncertainty and decision with certainty.

3 Managerial decision-making: information uncertainty reduction

There are two ways in which the value of information for the manager is significantly increased. The first way is obtaining the sufficient amount of information in time and with the content that the manager can use for their decision-making. The second way is increasing the precision of the estimates of the future values of quantities and of the output of processes occurring in economy. Every manager can make an intuitive estimate based on their experience by looking at the present and past development. These pragmatic estimates based on monitoring the previous process development offer valuable base information for decision-making. An estimate obtained this way is in the further text referred to as an expert estimate. Many expert estimates are made without any mathematical or other scientific model procedures or algorithms.

In the process of decision-making itself needs to be included a quantitative estimate of risk e.g. based on uncertainty, and also the calculation of effect/losses of risk reduction/increase. Such a process of risk reduction in managerial decision-making is represented in a diagram in Figure 1.

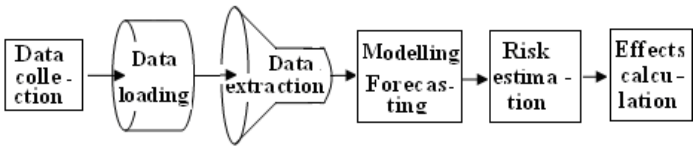


Figure 1 A diagram of uncertainty reduction in managerial decision-making

The first two blocks in Figure 1 represent activities connected with collecting and storing data. In company information systems such activities are carried out by tools known as ETT **(**Extraction, Transformation, Transport), ETL (Extraction, Transformation, Loading) tools. The relevant data extraction block is concerned with obtaining important data and the information about the relations among them. Such information is obtained

by statistical analyses known from descriptive statistics and with the support of graphic tools. It is also necessary to eliminate the data which are redundant for the given process.

In the modeling and forecast block, before making the decision itself, the manager must select a suitable forecasting model for determining the forecast. By selecting a suitable forecasting model according to the character of the monitored process, the manager can positively influence the quality of the forecast e.g. by increasing the precision of the prognosis.

Risk estimate block is important for comparison of the degree of how the attitude and the situation of the manager is changing during their decision-making. Risk change (reduction) affects the quality of the decision, i.e. uncertainty reduction must produce the decision effect. Numerical value of risk scene can be used for comparison of suitability of individual forecasting models or methods which are used to produce forecasts of future values. Different procedures were suggested for calculating uncertainty and thus also risk scene assessment. E.g. in [6]. the quantification of uncertainty in forecasting models is based on the analysis of variance forecast errors, in [4] the fuzzy set theory is used for calculation of forecast risks, etc. In the following chapter, the procedure for risk scene assessment on the basis of confidence intervals based on the probability is introduced.

The last block in Figure 1 is a block in which effect caused by uncertainty change (benefit or losses) is estimated. Economic quantities such as profit, turnover increase, cost savings or even economy in time are comprehensible quantities for every manager in every sphere of management. These quantities are used to compare individual alternatives of decisions. How the decision effect calculation will follow up the preceding forecast will depend on how costs are determined in a specific activity. The costs function will be different in solving tasks where e.g. stochastic inventory models are used, and it will be different in case of profit calculation in securities trade. A specific way of calculating effects of uncertainty reduction on practical example is given in [1].

4 Reducing uncertainty with the use of forecasting models

We will verify the sequence of steps for uncertainty calculation and reduction according to the diagram in Figure 1 by applying it to managerial decision-making at a transport company. Every month a transport company attends to a certain number of transport facilities according to the customer's requirements. It is the manager's task to forecast the number of the facilities and make sure that the company meets the monthly requirements of customers for the capacity of the transport facilities without delays.

4.1. Reducing Risk Scene of Managerial Decision-making in Attending to Transport Facilities

In the following section we will give an example of transport facilities number forecasting with ways of assessing risks and effects using forecasting models. On the basis of the obtained prognoses from these models, we will determine their prognostic precision, asses their entropy. In the next sub-chapter we will demonstrate the procedure of quantification of effects arising from the entropy reduction by using different forecasting models.

Managers of transport companies have at their disposal the time series of monthly observation of the numbers of transport facilities attended to in the period from 1990 to 2005, which comprises 192 observations. The development of these values in time is shown in Figure 2. Figure 2 shows that the observed values of the numbers of facilities attended to in the individual months do not prove any irregularities, jumps or periodical variation.

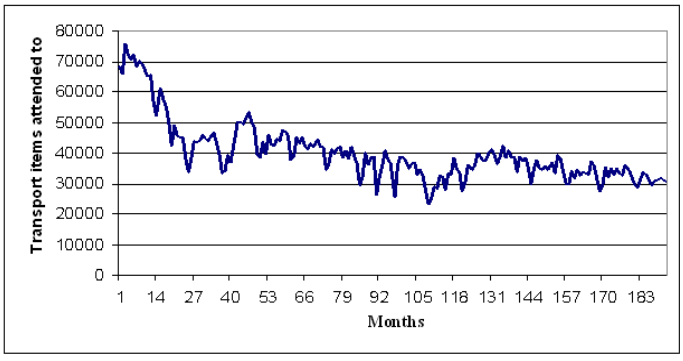


Figure 2 Monthly Numbers of Transport Facilities Attended to from 1999 to 2005

First, based on the development of the observed data, ex-post forecasts of the numbers of the transport facilities attended to in December 1999 and in December 2005 were made. Although these two values are in fact known, their estimates were made because the actual values will be used to compare the precision of the forecast. Forecast estimates were made in two ways for both months. One estimate was made by the manager (technologist) based on their experience from the past development of the values and the knowledge of the technological processes of attending to the facilities. This forecast is marked as an expert estimate. The second estimate was made by neural networks of the perceptron type [2] with the net determination based on the gradient method. The values of these estimates are given in Table 1. Table 1 demonstrates to what degree expert estimates and neural networks estimates approximate the actual values for December 1999 and December 2005.

December 1999		December 2005	
Actual value	33846	Actual Value	30621
*Expert estimate	41819	*Expert estimate	29845

*Expert estimates were made by Technical deputy master of the Slovak Railways

Table 1 The Actual Values and the Values of the Forecasts of the Number of the Facilities

Expert estimates of the numbers of the transport facilities attended to in the individual months in 1999 and in 2005 were then made by a technologist. Finally, estimates of prognoses for individual months in 1999 and in 2005 were made with the use of other forecasting models. The calculated MAPE (Mean of the Absolute Percentage Errors) values according to the individual forecasting models in 1999 and in 2005 are given in Table 2. Table 2 shows that the forecasting models based on artificial intelligence (the last three models in Table 2) achieve more precise results than the classical forecasting models based on the probability theory.

Year	Expert estimate	Regress. Analysis	Exponent. smoothing	Winter´s algorithm
1999	34.85	14.91	14.19	9.14
2005	6.02	5.27	6.63	4.99
Year	Direct smoothing	Adapt. Algorithm	GMDH algorithm	Neural network
1999	17.40	13.63	8.54	6.68
2005	18.23	4.23	4.15	3.99

Table 2 The Mean of the Absolute Percentage Error in the Forecasts with the Use of Forecasting Models in 1999 and in 2005 for the Next 12 Months in %

For the assessment of the estimate uncertainty degree, the method of confidence intervals for point forecasts was used. In this case it is possible to test the H_0 hypothesis of the expected type of probability distribution to determine confidence intervals provided that residuals have a normal probability distribution, and this hypothesis can be verified using χ^2 test of good fit on levels of significance set in advance. Using the χ^2 test of good fit, the H_0 hypothesis was verified on the level of significance $\alpha = 0.05$ and $\alpha = 0.01$, and this hypothesis claims that the residuals of the forecasted values from the actual values can be considered as a data file with a normal probability distribution. The confidence interval can be then calculated according to the following expression

$$x \in \left\langle \bar{x} - k_\alpha \,.\, \frac{\sigma}{\sqrt{n}},\ \bar{x} + k_\alpha \,.\, \frac{\sigma}{\sqrt{n}} \right\rangle \tag{1}$$

where k_α is the critical value of the standardized normal probability distribution, α is the level of significance, σ is the standard deviation, n is the number of observations, $\bar{x}$ is the expected value.

For the chosen probability $P = 0.95$, the confidence interval of the expert estimate will have the span $\langle 27352.27\,,\, 40339.73 \rangle$. This interval determines the limits which the expert estimate value will not exceed with 95% probability. The value $\alpha = 1 - P = 0.05$ is the so-called level of significance, which means the probability that a random variable of the expert estimate will acquire a value outside the interval $\langle 27352.27\,,\, 40339.73 \rangle$. Analogically, with the probability $P = 0.95$ was calculated the confidence interval for the expected value of the prognosis by the forecasting model based on neural networks with the values $\langle 31931.73,\ \ 35760.28 \rangle$.

Interesting about the support of the preference of forecasting models based on neural networks to manager´s expert estimates in managerial decision-making is the information about the probability change. The calculation of this probability is possible from expression (1) as the level of significance *k*

$$k = \overline{x} - a\frac{\sqrt{n-1}}{\sigma_{est}} \tag{2}$$

where α is the lower limit of the forecast interval of the prognosis calculated by neural network. E.g. in 1999 with the standard deviation σ_{est} = 10989.64 and expected value (mean) $\overline{x} = 33846$, which was calculated using estimates of prognoses for individual months in year 1999

$$k = 33846 - 31931{,}73\frac{\sqrt{12-1}}{10989.64} = 0.577.$$

According to the critical values of the standardized normal distribution, to $k_\alpha = 0.577$ appertains $\alpha = 0{,}57$. This implies that the probability that the mean value will fall into the narrower (more precise) interval will change from (1 – 0.577) = 0.423, i.e. from 42.3% to 95%. That is 52.7% growth.

4.2. Entropy as a Measure of Uncertainty

Another measure of uncertainty used in the theory of information is entropy [8]. Entropy and also uncertainty is expressed by the amount of information that we get after performing an experiment. For example, if we get a message that an event *A* has occurred with probability *P*(*A*), we also get information *I*(*A*) equal $-\log_2 P(A)$ bit. In case the event *A* consists of a finite amount of measured events, i.e. subsets of probabilistic space Ω while $A_i \in A$ for $i = 1,\ 2,...n$, $\Omega = \bigcup_{i=1}^{n} A_i$ and $A_i \bigcap A_j = 0$ for $i \neq j$ is valid, then the entropy expressed by Sannon´s definition is.

$$H(P) = \sum_{i=1}^{n} I(A_i).P(A_i) = -\sum_{i=1}^{n} I(A_i).\log_2 P(A_i) \tag{3}$$

In this connection, a very important question is, how will the entropy change if the estimate is more precise? The probability used in the relation for the calculation of entropy is the probability that the estimate value will fall into the narrower 95% confidence interval.

In case of an expert estimate in 1999, this probability is 45%. In case of the prognosis based on the forecasting model based on neural networks, this probability is 95%. Then

$H_{\text{expert estimate}}(P) = -\log_2 0.43 = 1.2176$ bit

$H_{\text{forecasting model}}(P) = -\log_2 0.95 = 0.074$ bit

By using the forecasting model, entropy in 1999 is reduced by 1.1436 bit. Analogically, the entropy values in 2005 are the following

$H_{\text{expert estimate}}(P) = -\log_2 0.83 = 0.26882$ bit

$H_{\text{forecasting model}}(P) = -\log_2 0.95 = 0.074$ bit.

By applying the forecasting model in 1999, the entropy value was reduced by 1.1436 bit, in 2005 by 0.19482 bit. In both cases, the application of the forecasting models led to entropy reduction, which makes it possible to make decisions with larger effect. Entropy reduction in 2005 is less substantial than in 1999. That is understandable given the more balanced and more regular development of the time series of the forecasted quantity in the last third of its development, as can be seen in Figure 2.

4.3. Uncertainty as the Standard Deviation

The standard deviation is used in literature as the degree of uncertainty and risk [5]. As far as relevancy is concerned, it is probably the easiest and, for managerial practice, the most comprehensible way of expressing and quantification of uncertainty. While the entropy indicated in the information unit bit is at present a still relatively abstract and almost non-used measure for expressing risk in the sphere of managerial decision-making. Uncertainty in the sense of the standard deviation has a higher informative value for managers. Uncertainty expressed by the standard deviation has one drawback, which is unit incompatibility. Entropy is indicated in bits. Despite this fact, as we could see in the given examples, it is easier to work with entropy as the standard deviation. It is possible to state that reduction of entropy of the forecast system was achieved when its standard deviation of forecast errors was reduced. It can be clearly seen in expression (1). In technical systems, rule 3 σ is used which in the figurative meaning provides information about which interval the forecast will almost certainly fall into.

Therefore, it provides certainty instead of uncertainty. But it is a certainty which will not push the manager forward with his decision-making if there is a big standard deviation. The real solution leading to the support of decision-making is reducing uncertainty of the forecast system by using a better forecasting model which will achieve lesser variability of prognosis errors. Described in [7], on the basis of prognosis errors analysis, is a method of searching for such a forecast horizon for which entropy and thus also prognosis risk is minimal.

5 Conclusion

In the present paper we showed the procedure of quantitative assessment of risk scene based on probability terms using confidence intervals for point estimates of economic quantities. We build upon measuring uncertainty based on information entropy indicated in bits and on measuring based on prognosis confidence interval, where uncertainty is expressed in terms of the span of the confidence interval and the probability that by using forecasting model the set prognosis limits around the expected value will not be exceeded. Both approaches to measuring uncertainty were assessed from the viewpoint of utilization in managerial decision-making using forecasting models based on an expert estimate, statistical models, and neural networks models.

The results of the study showed that there are more ways of approaching the issue of measuring risk in managerial decision-making in companies. It was also proved that it is possible to achieve significant risk reduction in managerial decision-making by applying modern forecasting models based on information technology such as neural networks developed within artificial intelligence.

Acknowledgements

This work was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070).

References

[1] Frano, M.: Lowering the Entropy in Managerial Decision Making Using Suitable Prediction Algorithms. Ph.D. Thesis, The University Press of Zilina, 2010.
[2] Hertz, J., Krogh, A., Palmer, R.G.: *Introduction to the Theory of Neural Computation*. Redwood City: Addison-Wesley, 1991.
[3] Huang, C.F.: A Note on the Difference between Disaster Prediction and Risk Assessment in Natural Disasters. New Perspectives on Risk Analysis and Crisis Response, **9**, Atlantis Press, Paris (2009), 1-7.
[4] Kahraman, C., and I. Kaya: Fuzzy process accuracy index to evaluate risk assessment of drought effects in Turkey. *Human and Ecological Risk Assessment*: An International Journal 15(4), (2009), 789–810.
[5] Marcek, D., Marcek, M., Babel, J.: Granular RBF NN Approach and Statistical Methods Applied to Modelling and Forecasting High Frequency Data. *Journal of Computational Intelligence Systems* 2-4, (2009), 353-364.
[6] Marcek, M., Pancikova, L., Marcek, D.*: Econometrics and Soft Computing.* The University Press of Zilina, 2008.
[7] Marcek, M., Petrucha, J., MarcekD.: Statistical and NN Forecasting Models of Financial Data: Making Inferences about the Accuracy and Risk Evaluation. *Journal of Mult.-Valued Logic & Soft Computing*, Vol. **1**7, (2011), 321-337.
[8] Sannon, C.E.: A Mathematical Theory of Communication. *Bell System Tech. J.*, **27**, (1948), 379-423.

Application of fundamental analysis methods to compare efficiency of complex portfolios consisting of values listed on stock exchange

Adrianna Mastalerz-Kodzis[1]

Abstract. The article presents theoretical basis and practical applications of selected quantity methods that can be used in building share fundamental portfolio, where elements of fundamental analyses and of classical portfolio theories are included. The new approach to creating a portfolio of securities, based on multidimensional comparative analysis is an alternative to models of Markowitz and Sharpe in portfolio analysis. This work shows theoretical basis of used methods and results of carried out empirical analyses.

Keywords: fundamental analysis, multidimensional comparative analysis, TMAI, BMS.

JEL Classification: C3, C8, G1, E4

AMS Classification: 91B28

1 Introduction

Due to changing market mechanisms and conditions, investing into financial markets is a compound process running is time and space. Uncertainty in obtaining investment goal, investment time, experience and psychological are the inseparable attributes in this process. However, the rational investor bases their decisions on intuition, experience and knowledge.

The aim of this work is to present models to support the investor in decision taking, which include new market tendencies. The process of investing into financial markets is a dynamic process depending on frequent changes, whose direction and impact is difficult to predict in the long periods of time.

2 Methodological elements of performed analyses.

Contemporary works on finances, econometrics and financial engineering present many methods which may help the investor in making decisions, among which we may distinguish methods of econometric predictions as well as methods of technical and fundamental analyses. In this article, selected methods taken from **fundamental analyses** and some notions of **classical portfolio theory** were used. The article consists of two basic parts. The first presents formulas and references as well as applied methods for data analyses; the other is of empiric character.

2.1 Multidimensional comparative analysis -TMAI construction.

When taking an investment decision on investment market, one should not base the decision only on such tools, which allow to analyses phenomena characterized by a single feature.

It is recommended to use methods that analyses complex phenomena, which are presented by a few (more than one) characteristics. **Multidimensional comparative analysis** (statistical comparative analysis) provide methods allowing to perform analysis of at least two variables, which describe the examined phenomenon. With such a device, we can compare different objects (companies or stock values), which are described by many features. On the basis of many data matrixes on objects, many taxonomic measurements can be built. The device may be also used to carry out an economic and financial check-up of a company. The estimation of *fundamental strength* of a given company by multidimensional comparative analysis is much more effective than the use of a single dimensional statistical method.

Fundamental portfolio of securities, built on the basis of multidimensional analysis is a long -term portfolio (is supposed to bring profit in long periods of time). Such a portfolio is highly efficient as it selects strong companies with regard to their economic and financial standing. Moreover, this portfolio is safe and stable in long periods of time.

The taxonomic measurement may be used to examine the attractiveness of investment, **fundamental power** of a company, which is its economic and financial condition. The analysis covers the period of three to five years

[1] Adrianna Mastalerz-Kodzis, Ph.D., University of Economics in Katowice, adamast@ue.katowice.pl

back. The taxonomic measurement, which allows to work out fundamental estimation of a company, is called **taxonomic measurement of investment attractiveness (TMAI)**, in which by one number it is possible to present the standing of a stock-exchange company. While building TMAI measurement, the matrix presents diagnostic features of a company financial standing. The point is to choose out of many data, the most important ones, which can be done in different ways depending on the analyst and access to data. A properly selected set of data leads to an accurate evaluation of the company financial condition (Jajuga [1], Jaworski [2]).

The stages of building a fundamental portfolio of securities are as follows:

1. Analysis of company: macroeconomic, sector, estimation of company financial standing, and estimation of share interior value.
2. Synthetic measurement construction TMAI to evaluate fundamental power of company.
3. Building the function of aim, responsible for fundamental measurement synthetic maximizing.
4. Presenting limiting conditions.
5. Optimal solutions.

Building a taxonomic measure consists of three stages (Łuniewska, Tarczyński [3]), (Tarczyński [5])]. Having data matrix, we normalize (standardize) the values, following the formula

$$z_{ij} = \left(x_{ij} - \bar{x}_j\right) \Big/ s_j \tag{1}$$

where:

$\bar{x}_j$ – mean of feature j,

s_j – standard deviation for feature j.

Next, the module method is used, and in the normalized matrix of m variables, the highest value is taken, module z_{0j}. The Euclidean distance from the module is calculated, using the formula

$$d_i = \sqrt{\sum_{i=1}^{m}\left(z_{ij} - z_{oj}\right)^2 \Big/ m} \tag{2}$$

The shorter the distance of the given object from the module, the lower is the value d_i. The obtained variable is not normalized, which next is transformed into a stimulant using the formula

$$z_i = 1 - \left(d_i / d_0\right) \tag{3}$$

where:

z_i – taxonomic development measure for object i,

d_i – distance of i object from module,

d_0 – standard to assure that variable z_i will take values ranging from 0 to 1, for example $d_0 = \bar{d} + 2s_d$, where $\bar{d}$ – mean d_i and s_d – standard deviation d_i.

In order to include weights in taxonomic measure, the formula (2) is to be modified in the following way

$$d_i = \sqrt{\sum_{i=1}^{m} w_i\left(z_{ij} - z_{oj}\right)^2} \tag{4}$$

where w_j are the values calculated according to formula $w_j = V_j \Big/ \sum_{i=1}^{m} V_i$, and $V_j = s_j \Big/ \bar{x}_j$ is the variability coefficient of j diagnostic variable for primary data (before normalization).

2.2 Non-norm synthetic measurement BMS

The analysis include the BMS measurement and following transformation formula of primary data transformation (Ostasiewicz [4])

$$y_{ij} = \left(x_{ij} - x_{\min,i}\right) \Big/ \left(x_{\max,i} - x_{\min,i}\right) \tag{5}$$

due to which variables become non-nominated quantities and receive the values ranging [0,1], retaining their different variance. The included synthetic measurement is the arithmetic mean of

$$BMS_i = \frac{1}{n}\sum_{j=1}^{m} \alpha_j y_{ij}\,, \tag{6}$$

where α_i equals -1, when the variable is destimulant and 1 when it is the stimulant; the higher the variable value, the higher the position in hierarchy is.

3 Inclusion of fundamental analyses selected elements and portfolio theory into the process of decision making.

Data taken from Stock Exchange in Warsaw and from London Stock Exchange were used for empiric analysis. The data related to companies dealing with the leading stock indices in the field of raw materials and output.

The selection of the data and indices showing fundamental strength of companies is subjective and depends on preferences of researcher and availability of data. Publication of data on fundamental strength of stock companies is different at various stock exchanges all over the world. Therefore, not always can you use the same model at different stock exchanges; the component elements of taxonomic measurements frequently have to be adjusted to published data.

Research period, namely years from 2009 to 2012 is described as period of financial crisis in the world, when investing into stock was very risky, and probably still is. Frequently, it was not connected with obtaining profit but rather with minimizing losses. Therefore, it is of primary importance to choose safe values, though there is no guaranteed profit.

3.1 Analysis on the basis of data taken from Stock Exchange in Warsaw

Empirical analysis includes data taken from Stock Exchange in Warsaw, namely data that characterize financial condition of companies creating WIG 20 index and those belonging to sectors of raw materials and fuels. Data was obtained from five companies in the period from 01.01 2009 to 31.12 2011. Values characterizing fundamental strength were selected, and in Table 1 (basis of data from: www.gpw.pl, www.bankier.pl, www.money.pl) averaged economic-financial indices covering the years were presented:

- Net profit margin index (net profit/net income from sales)
- ROA return of assets index (net profit/assets altogether)
- ROE return of equity index (net profit/own capital)
- Earning per share index (net profit/number of issued shares)
- Index P/BV (market price of share/company accountancy value for one share)

Company	*Net profit margin*	ROA (%)	ROE (%)	Earning per share (zł)	P/BV
BOGDANKA	17.5933	7.6867	11.0533	6.2967	1.7433
KGHM	32.4767	24.6933	34.0933	30.1933	1.77
LOTOS	3.99	4.32	10.2333	5.7333	0.5667
PGNIG	8.33	5.1867	7.6167	0.3033	0.9833
PKNORLEN	2.32667	3.7767	9.1533	4.71	0.77

Table 1 Results of empirical analyses, economical and financial company data.

The following values were calculated: historical rate of return R, standard deviation of return rate s, asymmetry coefficient A, and parameter β. TMAI were calculated for indices from Table 1 according to formula (1) – (4) and BMS values – according to formula (5) – (6)

Next, the indices were standardized using formula (1). Table 2 presents standardized values of variables as well as TMAI values. Module method was used (formula (2) and (3)). Every variable was given the highest value to build the module object. The distance of every variable from the module was calculated and the Euclidean distance was applied. The formula was modified by including weights based on variability coefficients (formula (4)).

Company	R	s(R)	A	β	BMS	TMAI
BOGDANKA	0.0010	0.0184	0.503941	0.6848	0.449926	0.488012
KGHM	0.0021	0.0298	-0.925854196	1.7266	0.707703	1
LOTOS	0.0012	0.0261	0.138393	1.0477	0.455379	0.353938
PGNIG	0.0003	0.0179	0.389455	0.6195	0.7165	0.348772
PKNORLEN	0.0006	0.0252	0.6377799	1.1321	0.402863	0.3392

Table 2 Results of empirical analyses for given companies. TMAI and BMS values.

Next, fundamental portfolio was built and with the use of Solver, the following optimization problem was solved. Calculation results are presented in Table 3. It turned out that basing on the above mentioned data, it is profitable to invest into companies like: BOGDANKA, KGHM, PNNIG . Adding the condition of including parameter value β did not change optimal result (task 2 and 3 as well as 5 and 6). Therefore, if one invests, according to model taken from task 1, dated 04.01.2010 the amount of PLN 1.000.000, the rate of return on

23.04.2012 is 46.19%. However, the investment horizon plays here an essential role as if one invested the same amount one year later, on 03.01.2011, the loss would be 5.4 %.

Model 1	Model 2	Model 3
$f = \sum_{i=1}^{5} TMAI_i x_i \to \max$ $\sum_{i=1}^{5} R_i x_i \geq \overline{R}$ $\sum_{i=1}^{5} s_i x_i \leq \overline{s}$ $\sum_{i=1}^{5} x_i = 1$ $x_i \geq 0 \quad i = 1,\ldots,5$	$f = \sum_{i=1}^{5} TMAI_i x_i \to \max$ $\sum_{i=1}^{5} R_i x_i \geq \overline{R}$ $\sum_{i=1}^{5} s_i x_i \leq \overline{s}$ $\sum_{i=1}^{5} A_i x_i \geq \overline{A}$ $\sum_{i=1}^{5} x_i = 1$ $x_i \geq 0 \quad i = 1,\ldots,5$	$f = \sum_{i=1}^{5} TMAI_i x_i \to \max$ $\sum_{i=1}^{5} R_i x_i \geq \overline{R}$ $\sum_{i=1}^{5} s_i x_i \leq \overline{s}$ $\sum_{i=1}^{5} A_i x_i \geq \overline{A}$ $\sum_{i=1}^{5} \beta_i x_i \leq \overline{\beta}$ $\sum_{i=1}^{5} x_i = 1$ $x_i \geq 0 \quad i = 1,\ldots,5$

Model 4	Model 5	Model 6
$f = \sum_{i=1}^{5} BMS_i x_i \to \max$ $\sum_{i=1}^{5} R_i x_i \geq \overline{R}$ $\sum_{i=1}^{5} s_i x_i \leq \overline{s}$ $\sum_{i=1}^{5} x_i = 1$ $x_i \geq 0 \quad i = 1,\ldots,5$	$f = \sum_{i=1}^{5} BMS_i x_i \to \max$ $\sum_{i=1}^{5} R_i x_i \geq \overline{R}$ $\sum_{i=1}^{5} s_i x_i \leq \overline{s}$ $\sum_{i=1}^{5} A_i x_i \geq \overline{A}$ $\sum_{i=1}^{5} x_i = 1$ $x_i \geq 0 \quad i = 1,\ldots,5$	$f = \sum_{i=1}^{5} BMS_i x_i \to \max$ $\sum_{i=1}^{5} R_i x_i \geq \overline{R}$ $\sum_{i=1}^{5} s_i x_i \leq \overline{s}$ $\sum_{i=1}^{5} A_i x_i \geq \overline{A}$ $\sum_{i=1}^{5} \beta_i x_i \leq \overline{\beta}$ $\sum_{i=1}^{5} x_i = 1$ $x_i \geq 0 \quad i = 1,\ldots,5$

where:

$TMAI_i$ – taxonomic measure of investment attractiveness for *i*-company,

BMS_i – synthetic measure of investment attractiveness for *i*-company,

x_i – contribution of *i*-share in portfolio,

$\overline{R}$ – average rate of return for companies,

$\overline{s}$ – mean standard deviation,

$\overline{A}$ – mean asymmetry coefficient,

$\overline{\beta}$ – Mean coefficient β.

Such analyses, based on the a/m formula may be carried out for stock values from different stock sectors, and also take advantage of other data describing fundamental strength of companies. In this way, it is possible to compare portfolios, to which belong companies from given sectors (possible comparisons of companies), or choosing values from different sectors, having favourable qualities relating to selected indices of fundamental analysis.

Solution	Model 1	Model 2 Model 3	Model 4	Model 5 Model 6
X1=	0.555388292	0.751574064	0	0.468692832
X2=	0.444611708	0.248425821	0.46735873	0.223803519
X3=	0	0	0	0
X4=	0	1.14676E-07	0.532641269	0.307503649
X5=	0	0	0	0
Rate of return (%) 4.01.2010- 23.04.2012	46.1892348	54.4636349	14.423967	37.7173163
Rate of return (%) 3.01.2011- 23.04.2012	-5.3715624	0.37761	-3.16010934	2.7590323

Table 3 Fundamental portfolio.

3.2 Analysis based on data taken from London Stock Exchange

Research was based on data included in FTSE 350 High Field index and those belonging to Mining and Oil & Gas Producers sector. Values (table 4) informing about financial condition of companies on the basis of data presented on site www.londonstockexchange.com ,were selected. Averaged data for period from 2009 to 2011.

Company	Net Asset Value per Share	Profit After Tax	Return on Capital Employed	Earnings per Share - Basic	Equity Holders of Parent Company
BP	414.6666667	13177.33333	0.127467	68	12853
RDS A	3022	21459	0.156167	343.3333333	21187.66667
RDS B	3593.666667	21459	0.156167	343.3333333	21187.66667
AVOCET MINING	159.3333333	-1.19666667	0.018433	16.33333333	35.00666667
NEW WORLD RESOURCES	208.3333333	95.78666667	0.123667	1446.44	100.1266667

Table 4 Economical and financial company data, averaged for period from 2009 to 2011.

Synthetic measure values TMAI and BMS were calculated according to formula (1) – (4).

Company	Performance P	Volatility V	TMAI	BMS
BP	9.26%	2.521	0.622379	0.402399
RDS A	3.92%	1.468	0.934997	0.428834
RDS B	3.75%	1.533	1	0.434159
AVOCET MINING	5.60%	2.964	0.290343	0.268459
NEW WORLD RESOURCES	3.33%	3.64	0.259028	0.255849

Table 5 P, V, TMAI and BMS values for companies.

Next, fundamental portfolios 1 and 2 were built.

Model 1	Model 2
$f = \sum_{i=1}^{5} TMAI_i x_i \to \max$	$f = \sum_{i=1}^{5} BMS_i x_i \to \max$
$\sum_{i=1}^{5} P_i x_i \geq \overline{P}$	$\sum_{i=1}^{5} P_i x_i \geq \overline{P}$
$\sum_{i=1}^{5} V_i x_i \leq \overline{V}$	$\sum_{i=1}^{5} V_i x_i \leq \overline{V}$
$\sum_{i=1}^{5} x_i = 1$	$\sum_{i=1}^{5} x_i = 1$
$x_i \geq 0 \quad i = 1,\ldots,5$	$x_i \geq 0 \quad i = 1,\ldots,5$

where:

$TMAI_i$ – taxonomic measure of investment attractiveness for *i*-company,
BMS_i – synthetic measure of investment attractiveness for *i*-company.
x_i – contribution of *i*-share in portfolio,
$\overline{P}$ – mean value of P,
$\overline{V}$ – mean value of V.

With the use of Solver, the optimum task was solved.

SOLUTION	MODEL 1, MODEL 2
X1=	0.258076
X2=	0
X3=	0.741924
X4=	0
X5=	0
Rate of return (%)	
4.01.2010- 23.04.2012	-1.472553484

Table 6 Fundamental portfolio.

Optimal solution is characterized by negative rate of return, which is connected with big fluctuations of stock value prices during crisis times.

4 Conclusion

Investor, making a decision, is guided by the idea of obtaining highest possible profit at the lowest possible risk. In order to make the optimal decision, a rational investor is supported by different methods enabling analyses. However, the investor takes risks relating to making the right choices of analytical tools used for describing the task and selecting data for building a model. Therefore, the conclusion to be drawn is to include many methods and data for performing analyses, which allow to evaluate the examined object, taking into consideration as many features as possible.

References

[1] Jajuga K., Jajuga T., *Inwestycje.* PWN, Warszawa 2006.
[2] Jaworski J. *Teoria i praktyka zarządzania finansami przedsiębiorstw*. Cedetu, Warszawa 2010.
[3] Łuniewska M., Tarczyński W. *Metody wielowymiarowej analizy porównawczej na rynku kapitałowym*, PWN, Warszawa 2006.
[4] Ostasiewicz W. (Red.) *Statystyczne metody analizy danych.* Wydawnictwo AE we Wrocławiu, Wrocław 1999.
[5] Tarczyński W. *Fundamentalny portfel papierów wartościowych*. PWE, Warszawa 2002.

Web sites

http://www.bankier.pl
http://www.gielda.onet.pl
http://www.gpw.pl
http://www.londonstockexchange.com
http://www.money.pl

The Ordinal Consensus Ranking Problem with Uncertain Rankings

Jiří Mazurek[1]

Abstract: In the ordinal consensus ranking problem (OCRP) a set of k decision makers rank a set of n alternatives with regard to one overall criterion (or a set of criteria) from the 1st place to the n^{th} place. The goal is to find a consensus ranking expressing an opinion of a group. The aim of this article is to present a model for OCRP solution with uncertain rankings. This approach is more suitable than classic approach with certain rankings, as the latter case doesn't allow for imprecise information or uncertainty often involved in real decision-making processes. In this paper uncertain ranking g_{ij} is defined as a decision maker's confidence that an alternative i is ranked at the j^{th} position, where $g_{ij} \in [0,1]$ and $\sum_j g_{ij} = 1$ for all i. In the model for OCRP solution with uncertain rankings generalized means operator is used for ranking aggregation and the final consensus ranking is obtained by the use of a binary dominance relation. The model can be extended to multiple criteria or different weights of decision makers, and it can handle the cases with certain rankings as well. Also, the model's setting enables to evaluate decision makers' preferences in terms of inconsistence and indecisiveness.

Keywords: group decision making, ordinal consensus ranking problem, preference ranking, uncertain ranking.

JEL Classification: D71
AMS Classification: 90B50

1 Introduction

The ordinal consensus ranking problem (OCRP) represents a special case of (multi-criteria) group decision making which history dates back to the works by Borda [2] and Condorcet [3] from the late 18th century. In OCRP a set of decision makers (experts) rank a set of alternatives with regard to a given set of criteria or one overall criterion. The goal is to find a consensus ranking expressing an opinion of a group. In general, there are two different classes of (classic) methods for OCRP solution, ad-hoc methods and distance based, which are briefly discussed in the next Section. The 'state of the art' of the ordinal consensus rankings problem can be found in [4]. Recently, some new methods of solution were proposed (see [11, 13]) and research has focused on examination of conditions, under which the same result is obtained by different methods ([6, 7, 8, 12, 13]).

The aim of this article is to present a model for a solution to ORCP when alternatives' rankings are uncertain. This approach may be more suitable when compared to the classic approach with certain rankings, as the latter case doesn't allow for imprecise information or uncertainty often involved in real decision making processes. In OCRP it is assumed that each decision maker provides rank order (ranking) of all alternatives. However, in many cases a decision maker is not able to do so due to lack of knowledge, time pressure, imprecise information, etc. In such situations a decision maker can express his preferences in a form of ranking with some degree of uncertainty. If a decision maker is certain about his ranking, he assigns each alternative its position with full confidence (with the probability equal to 1). But if he is uncertain, he can give a degree of confidence that an alternative is placed at the n^{th} position by the number from [0,1].

The proposed model with uncertain rankings can be extended for multiple criteria and different weights of decision makers, and it incorporates cases with crisp rankings as well, thus providing a generalization to classic methods for OCRP solution. Moreover, model's setting enables to evaluate 'quality' of decision makers' preferences in terms of indecisiveness and inconsistence.

The paper is organized as follows: the classic approach to OCRP is discussed in Section 2; the proposed model with uncertain rankings is described in Section 3 followed by an illustrative example in Section 4. Model's extensions are presented in Sections 5 and 6, and Conclusions close the article.

[1] Silesian University in Opava, School of Business Administration in Karviná, Department of Mathematical Methods in Economics, mazurek@opf.slu.cz.

2 Methods for OCRP solution with certain rankings

In OCRP with certain rankings each decision maker (DM) ranks alternatives from the best to the worst. Each ranking constitutes total order on a set of alternatives, as each alternative is assigned precisely one position. The final group ranking (consensus) is achieved by some of these classic single criterion methods:

- **Ad-hoc methods**

-*Borda-Kendall's method of marks* (see [2] and [9]): each alternative is given a number of points corresponding to its rank. The best alternative is the alternative with the lowest total count (mark) or the lowest average.
-*Condorcet's simple majority rule* (see [3]): the best alternative is an alternative preferred over all other alternatives in pair-wise comparisons.
-*Maximize agreement heuristic* (MAH) by Beck and Lin [1]: all alternatives are pair-wise compared by all DMs and then ranked in the descending order according to their total number of preferences (*P*) or non-preferences (*N*).

- **Distance based methods**

In distance based methods rankings of DMs are converted into a vector, object-to-object or object-to-rank matrix (alternatives are displayed in rows and their position in columns) representation subsequently. Then, by the use of a suitable distance function on vector or matrix spaces, the consensus is searched through the space of all possible rankings (permutations of the order *n*), where the consensus is defined as the ranking with the minimal sum of distances to rankings of all DMs. Usually, the following l_l metric is used as a distance function:

$$d(A,B) = \sum_{i=1}^{n} \sum_{j=1}^{n} \left| a_{ij} - b_{ij} \right|,$$

where A (a_{ij}) and B (b_{ij}) are square matrices of the order n. Distance based methods include e.g. *Consensus ranking model* (*CRM*) by Cook and Kress [5] and *Distance-based ideal-seeking consensus ranking model* (*DCM*) by Tavana et al. [13].

However, all aforementioned methods share several limitations. They cannot handle ties or non-preferences between alternatives; they don't allow expressing a degree of preference among alternatives; they don't enable to express decision makers' importance (weight) and finally they assume precise information in the form of certain ranking of alternatives is provided by DMs. Above mentioned disadvantages can be put aside by the use of the proposed model with uncertain rankings.

3 OCRP with uncertain rankings

3.1 Uncertain rankings

In the context of this paper certain rankings (briefly c-rankings) have to be distinguished from uncertain rankings (u-rankings). C-ranking is represented by a binary preference matrix with rows corresponding to alternatives and columns to positions (for an example see Figure 1). These matrices are bistochastic, as there is precisely one 1 on each row and column. By analogy, u-ranking can be represented by a row stochastic matrix with elements in the interval $[0,1]$, see Figure 1. Formally, u-rankings are introduced by the following Definition 1.

$$K = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \qquad L = \begin{pmatrix} 0.3 & 0.3 & 0.2 & 0.2 \\ 0.2 & 0.4 & 0.2 & 0.2 \\ 0.4 & 0.3 & 0.2 & 0.1 \\ 0.1 & 0.2 & 0.4 & 0.3 \end{pmatrix}$$

Figure 1 Matrix representation of c-rankings (K) and u-rankings (L) of four alternatives A, B, C, and D. Matrix K gives following crisp ranking: (A, B, D, C)

Definition 1. *Let* $g_{ij}^{k} \in [0,1]$ *be confidence of the* k^{th} *decision maker that an alternative i occupies the* j^{th} *position, where* $i, j \in \{1,2,...,n\}$ *and* $k \in \{1,2,...,K\}$. *Then* $g_{ij}^{k} \in [0,1]$ *is called uncertain ranking (u-ranking), if:*

$$\sum_{j=1}^{n} g_{ij}^{k} = 1 \text{ for all } i \text{ and } k \qquad (1)$$

According to this definition u-rankings are normalized. Ordering of alternatives is uncertain (fuzzy) in the sense that each alternative can occupy each position, but generally with the different degree of confidence (see Table 1). Hence, uncertain rankings can be regarded as fuzzy measures on the set of positions. From a probability point of view, u-ranking of a given alternative can be interpreted as a probability mass function that assigns a value $p(x_i)$ to each position x_i, $i \in \{1, 2, ..., N\}$, such that:

$$p(x_i) \geq 0 \text{ and } \sum_{i=1}^{N} p(x_i) = 1 \tag{2}$$

3.2 Aggregation of uncertain rankings

DMs' u-rankings g_{ij}^{k} *ha*ve to be aggregated by aggregation functions or operators. In this paper generalized means [16] $h : [0,1]^n \to [0,1]$ are used in the following form:

$$h_{ij}(g_{ij}^{(k)}, \alpha) = \left[\frac{\sum_{k=1}^{K} g_{ij}^{(k)\alpha}}{K} \right]^{1/\alpha}, \tag{3}$$

where $\alpha \in R - \{0\}$. For $\alpha = 1$ we obtain the arithmetic mean, for $\alpha \to 0$ the geometric mean and for $\alpha = -1$ the harmonic mean [10]. In the model the averaging operator (3) with $\alpha = 1$ and $\alpha \to 0$ is used. The average (group) u-ranking of an alternative *i* at the position *j* is denoted as h_{ij}.

3.3 Ordering of alternatives

As the aim of OCRP is to establish the final group consensus ranking, alternatives must be compared and ordered finally. For a comparison of alternatives the following binary dominance relation is introduced:

Definition 2. *Let h_{ij} be the group u-ranking of the alternative i at the position j. Then, the cumulative group u-ranking H_{ij} of alternative i from the 1st to the jth position is given as:*

$$H_{ij} = \sum_{k=1}^{j} h_{ik} \tag{4}$$

Definition 3. *An alternative r dominates an alternative s $(r \succ s)$ if cumulative group u-rankings H_{rj} of an alternative r are at least equal to cumulative group u-rankings H_{sj} of an alternative s for every position j, and there is a position p such that cumulative group u-ranking H_{rp} is higher than H_{sp}:*

$$r \succ s \Leftrightarrow H_{rp} \geq H_{sp}, \forall n \in \{1, 2, ..., N\} \wedge H_{rp} > H_{sp}, \qquad 1 \leq p \leq N \tag{5}$$

The dominance relation (5) provides a partial quasi-order on the set of alternatives, as some alternatives might not be comparable and thus the final consensus ranking (and the best alternative) might not be unique.

3.4 The model

The proposed model for the solution of the ordinal consensus ranking problem with u-rankings is composed of three parts: u-rankings of individual decision makers, the aggregation (averaging) operator and the dominance relation.

Decision makers' uncertain rankings of each alternative, preferably in a matrix format, represent the model's input. The output of the model is the best alternative (or alternatives). The model proceeds in five steps:

1. Each DM gives u-ranking g_{ij} for each alternative according to his knowledge and confidence.
2. DMs' u-rankings are aggregated for each alternative and each position by the averaging operator (3). When other than arithmetic mean is used for aggregation, average group u-rankings have to be normalized subsequently.
3. For each alternative *i* cumulative group u-rankings H_{ij} is evaluated by (4). Because of normalization $H_{in} = 1$ for each alternative *i*.
4. All alternatives are pair-wise compared with the use of the dominance relation (5).
5. Alternatives are ranked according to their dominance.

It is possible to integrate an additional step between steps 1 and 2 evaluating decision makers' rankings in terms of indecisiveness and inconsistence (see Section 6 for details).

As certain rankings constitute only a special case (a subset) of uncertain rankings, they can be handled by the model as well. The next section illustrates how the model works.

4 Illustrative example

Four decision makers (DM_1 to DM_4) rank four alternatives A, B, C and D from the best to the worst. U-rankings of DMs are presented in Table 1. Rankings for each alternative are averaged with respect to DMs via relation (6) with $n = 4$, $K = 4$ and $\alpha = 1$ and they are shown in Table 2. Cumulative group u-rankings of all alternatives are presented in Table 3.

As for alternatives' comparison, from the dominance relation (7) we get:

$$A \succ B,\ A \succ C,\ A \succ D,\ B \succ C,\ B \succ D$$

Alternatives *C* and *D* are non-comparable. Therefore, we obtain two final rank orders: (A, B, C, D) and (A, B, D, C). In both cases the best alternative is A.

If the geometric mean ($\alpha \to 0$) is used for rankings aggregation instead of the arithmetic mean, results wouldn't change (see Table 4 and Table 5). However, the geometric mean is not appropriate operator for aggregation of rankings consisting of many 0.

DM_1	1st	2nd	3rd	4th	DM_2	1st	2nd	3rd	4th
A	0.4	0.2	0.1	0.3	A	0.5	0.3	0.1	0.1
B	0.3	0.3	0.2	0.2	B	0.3	0.3	0.2	0.2
C	0.1	0.2	0.4	0.3	C	0.2	0.3	0.4	0.1
D	0.1	0.4	0.4	0.1	D	0.1	0.3	0.5	0.1
DM_3	1st	2nd	3rd	4th	DM_4	1st	2nd	3rd	4th
A	0.6	0.2	0.1	0.1	A	0.4	0.4	0.2	0
B	0.3	0.4	0.2	0.1	B	0.3	0.3	0.3	0.1
C	0.2	0.2	0.3	0.3	C	0.1	0.2	0.4	0.3
D	0.1	0.2	0.3	0.4	D	0.1	0.3	0.4	0.2

Table 1 U-rankings of decision makers DM_1 - DM_4 for alternatives A, B, C, D

Alternative	1st	2nd	3rd	4th
A	0.475	0.275	0.125	0.125
B	0.3	0.325	0.225	0.15
C	0.5	0.225	0.375	0.25
D	0.1	0.3	0.4	0.2

Table 2 Group u-rankings of alternatives A, B, C, D for the 1st, 2nd, 3rd and 4th place

Alternative	1st	2nd	3rd	4th
A	0.475	0.75	0.875	1
B	0.3	0.625	0.85	1
C	0.15	0.375	0.75	1
D	0.1	0.4	0.8	1

Table 3 Cumulative group u- rankings of alternatives A, B, C, D for all places

Alternative	1st	2nd	3rd	4th
A	0.564	0.3	0.136	0
B	0.305	0.327	0.225	0.144
C	0.147	0.23	0.387	0.237
D	0.105	0.306	0.413	0.176

Table 4 The geometric mean of uncertain rankings of alternatives A, B, C and D

Alternative	1st	2nd	3rd	4th
A	0.564	0.864	1	1
B	0.305	0.632	0.857	1
C	0.147	0.377	0.764	1
D	0.105	0.411	0.824	1

Table 5 Cumulative group u-rankings of alternatives A, B, C and D

5 Extensions

The model's setting presented in Section 3 enables straightforward extensions in terms of decision makers' weights and multiple criteria:

- To each decision maker weights w_i can be assigned according to his/her importance or knowledge. For the aggregation of preferences, e.g. the weighted arithmetic mean can be used:

$$h_{ij}\left(g_{ij}^{(k)}, w_k\right) = \left(\sum_{k=1}^{K} g_{ij}^{(k)} \cdot w_k\right) / \sum_{k=1}^{K} w_k \tag{6}$$

- Alternatives can be ranked by more than one criterion, and in this case criteria themselves can be ranked in order of importance in the same way as alternatives. The overall u-ranking of each alternative is obtained by the aggregation over criteria of each DM and then over all decision makers (or vice versa) with the use of (6). Again, u-rankings should be normalized in the process.

6 The evaluation of decision makers' preferences

The model's framework allows evaluating experts' decisions in terms of *indecisiveness* and *inconsistence*. An expert is absolutely decisive, when he assigns each alternative value 1 for a given position and value 0 to all other positions, and indecisive otherwise. To evaluate indecisiveness, Shannon's entropy as a measure of uncertainty can be used [10]:

$$H\left(p\left(x\right)\right) = -\sum_{i=1}^{N} p\left(x_i\right) \log_2 p\left(x_i\right), \tag{7}$$

where $p(x_i)$ are probabilities assigned to values x_i, $i \in \{1, 2, ...N\}$; and $H\left(p\left(x_i\right)\right) = 0$ for $p\left(x_i\right) = 0$.

A decision maker is absolutely indecisive, if he provides u-rankings with the uniform distribution $p\left(x_i\right) = \frac{1}{N}$, $i \in \{1, 2, ...N\}$ for a given alternative (see an example on the left-hand side of Table 6). In this case, the entropy (7) is equal to the Hartley's information I (N) (Hartley's measure of *nonspecifity*) $I\left(N\right) = \log_2 N$. Because each decision maker provides u-rankings of N alternatives, DM's maximum indecisiveness IND_{max} is given as:

$$IND_{\max} = N \log_2 N \tag{8}$$

The overall DM's indecisiveness *IND* is given as:

$$IND = -\sum_{i=1}^{N} \sum_{j=1}^{N} g_{ij} \cdot \log_2\left(g_{ij}\right) \tag{9}$$

A DM is absolutely consistent in his judgment, if his sum of u-rankings for each position over all alternatives is 1, and inconsistent otherwise. Therefore, inconsistence *IND* in the model's setting is given as:

$$INC = \sum_{j=1}^{N} \left| \sum_{i=1}^{N} g_{ij} - 1 \right| \tag{10}$$

Maximum inconsistence $INC_{\max}$ is achieved when a DM assigns value 1 to the same position for all alternatives (see an example on the right-hand side of Table 6). Then from (10) we obtain:

$$INC_{\max} = 2\left(N - 1\right) \tag{11}$$

Relations (8-9) and (10-11) allow expressing the relative indecisiveness IND_r and reative inconsistence INC_r:

$$IND_r = \frac{IND}{IND_{\max}} \tag{12}$$

$$INC_r = \frac{INC}{INC_{\max}} \tag{13}$$

Extreme cases of experts' decisions are illustrated in Table 6. DM_1 is absolutely indecisive (*IND* = 2), but he is absolutely consistent (*INC* = 0), while DM_2 is absolutely decisive (*IND* = 0), but he is absolutely inconsistent (*INC* = 6). Unlike many other models, which pay little or no attention to the quality of experts' decisions, in the presented model experts' decisions can be easily and clearly scrutinized in terms of indecisiveness and inconsistence, and highly indecisive and/or highly inconsistent experts (such as DMs shown in Table 6) might be given lower weights or even may be excluded from a decision making process.

DM_1	1st	2nd	3rd	4th	DM_2	1st	2nd	3rd	4th
A	0.25	0.25	0.25	0.25	A	0	1	0	0
B	0.25	0.25	0.25	0.25	B	0	1	0	0
C	0.25	0.25	0.25	0.25	C	0	1	0	0
D	0.25	0.25	0.25	0.25	D	0	1	0	0

Table 6 DM_1 assigns each alternative and each position the same value 0.25, hence he is absolutely indecisive. DM_2 ranks all alternatives in 2nd position, and hence he is absolutely inconsistent

7 Conclusions

The aim of the article was to present a simple model for ordinal consensus ranking problem with uncertain rankings, and to illustrate the use of the model by examples. The model is more realistic for a solution of real-world problems involving uncertainty and imprecise information. Other advantages of the model include computational simplicity and extensions to multiple-criteria or different weights of decision makers. Moreover, in the model's setting experts' judgments can be evaluated in terms of indecisiveness and inconsistence. As certain rankings constitute the subset of uncertain rankings, the model provides generalization to classic methods for the ordinal consensus ranking problem solution.

Acknowledgements
The paper was supported by the Grant Agency of the Czech Republic (no. 402090405).

References:

[1] Beck, M. P., Lin, B. W.: Some heuristics for the consensus ranking problem. *Computers and Operations Research* **10** (1) (1983), 1–7.
[2] Borda, J. C. M´emoire sur les ´elections au scrutin. *Histoire de l'Acad´emie Royale des Sciences*. Paris, 1781.
[3] Condorcet, M. J. A. N.: *Essai sur L´Application de L´Analyse a la Probabilite des Decisions Rendues a la Pluralite des Voix*. Paris, 1785.
[4] Cook, W. D.: Distance-Based and Ad Hoc Consensus Models in Ordinal Preference Ranking. *European Journal of Operational Research* **172** (2006), 369-385.
[5] Cook, W. D., Kres, M.: Ordinal ranking with intensity of preference. *Management Science* **31** (1) (1985), 26-32.
[6] Cook, W. D., Seiford, L., Warner, S.: Preference ranking models: Conditions for equivalence, *Journal of Mathematical Sociology* **9** (1983), 125-127.
[7] Fiedor, J., Mazurek, J.: The Evaluation of the Entropy of Decision Makers' Preferences in Ordinal Consensus Ranking Problem. In: *Proceedings of the 29th International Conference on Mathematical Methods in Economics 2011* (Dlouhý, M., Skočdopolová, V., eds.). Professional Publishing, Prague, 2011, 157-162.
[8] Fishburn, P. C.: A comparative analysis of group decision methods, *Behavioral Science* **16** (1971), 538-544.
[9] Kendall, M.: *Rank Correlation Methods*, Hafner, New York, NY, USA, 3rd edition, 1962.
[10] Klir, G., Folger, T. A.: *Fuzzy Sets, Uncertainty, and Information*. State University of New York, 1988.
[11] Mazurek, J.: The Fuzzy Permutation Model for the Ordinal Ranking Problem Solution. In: *Proceedings of the 29th International Conference on Mathematical Methods in Economics 2011* (Dlouhý, M., Skočdopolová, V., eds.). Professional Publishing, Prague, 2011, 462-467.
[12] Richelson, J.: A comparative analysis of social choice functions III, *Behavorial Science* **3** (1978), 169-176.
[13] Tavana, M., Lopinto, F., Smither, J. W.: A Hybrid Distance-Based Ideal-Seeking Consensus Ranking Model. *Journal of Applied Mathematics and Decision Sciences*, 2007, 18 pages.

Modeling dependence and feedback in ANP with fuzzy cognitive maps

Jiří Mazurek[1], Zuzana Kiszová[2]

Abstract: The objective of multi-criteria decision making is to select the best alternative from a set of feasible alternatives with regard to a given set of criteria. To include dependence and feedback among criteria analytic network process (ANP) was proposed by T. L. Saaty. However, ANP have two disadvantages: firstly, it is difficult to provide correct network structure among criteria even for experts, and different structures lead to different results. Secondly, to form a supermatrix all criteria have to be pair-wise compared with regard to all other criteria, which is difficult and also unnatural. To circumvent these problems criteria network structure in ANP can be modeled with fuzzy cognitive maps. The aim of this article is to propose the hybrid eigenvalue-fuzzy cognitive map method (HEFCM) for the derivation of criteria weights. In the first step of HEFCM initial (local) weights of criteria are determined by Saaty's eigenvalue method. In the second step contribution to criteria weights from dependence and feedback is established by fuzzy cognitive map approach. Final (global) weights of all criteria are obtained by an aggregation of both weights. The proposed method is illustrated by an example.

Keywords: Analytic hierarchy process (AHP), analytic network process (ANP), criteria weights, decision making, fuzzy cognitive maps

JEL Classification: C02, C44
AMS Classification: 90B50, 91B06, 15A16, 15A18, 15B51

1 Introduction

The objective of multi-criteria decision making (MCDM) is to select the best alternative or object from a set of feasible alternatives or objects with regard to a given set of criteria. However, in real-world MCDM problems criteria (and other elements) are not independent as they influence each other. To include dependence and feedback into consideration analytic network process by T. L. Saaty was proposed (see [4, 5]).

However, ANP has two disadvantages [6]: firstly, it is difficult to provide a correct network structure even for experts, and different structures lead to different results. Secondly, to form a supermatrix all criteria have to be pair-wise compared with regard to all other criteria, which is also difficult and somewhat unnatural, as we ask themselves questions of the type: "How much is a criterion A more important than a criterion B with regard to a criterion C?" To circumvent these two problems criterias' network structure in ANP can be modeled with fuzzy cognitive maps.

Fuzzy cognitive maps are graphical tools introduced by B. Kosko [2] enabling to express dependence and feedback among concepts with different intensity given by a real number from [–1,1] or [0,1] interval. Fuzzy cognitive maps were found useful in many decision making areas such as politics, management, environmental protection or medicine.

The aim of this article is to propose and illustrate the use of the hybrid eigenvalue-fuzzy cognitive map method (HEFCM) for the derivation of global weights of criteria under dependencies and feedbacks based on Saaty's eigenvalue method and fuzzy cognitive maps, which can circumvent two disadvantages of ANP mentioned before. This new approach is compared to ANP in an illustrative example.

The paper is organized as follows: in section 2 AHP/ANP is briefly described, in section 3 fuzzy cognitive maps are introduced, section 4 provides description of our HEFCM method and in section 5 a numerical example is provided. Conclusions close the article.

[1] Silesian University in Opava, School of Business Administration in Karviná, Department of Mathematical Methods in Economics, mazurek@opf.slu.cz.

[2] Silesian University in Opava, School of Business Administration in Karviná, Department of Mathematical Methods in Economics, kiszova@opf.slu.cz.

2 Analytic hierarchy process (AHP) and analytic network process (ANP)

In AHP hierarchical structure of elements such as goal, criteria (sub-criteria) or alternatives is considered, where elements from higher levels of hierarchy influence elements from (immediately) lower levels, but not vice versa, and elements on the same level are considered independent. In this paper we limit ourselves to the classic 3-level hierarchy (goal, criteria and alternatives).

The basis of AHP/ANP is a pair-wise comparison of elements: the relative importance of elements from a given level of hierarchy with regard to an element on immediately higher level is expressed by a number on Saaty's fundamental scale (see Table 1). To derive weights of criteria, the pair-wise comparison matrix S (s_{ij}) is constructed, where $s_{ij} = v_i / v_j$ is the ratio of importance of an element i compared to an element j, $s_{ij} \in \{1,2,3,4,5,6,7,8,9\}$. Because $s_{ij} = 1/s_{ji}$ for all i and j, the matrix S is reciprocal (see Figure 1).

Weights of criteria (w) are determined by a principal eigenvector belonging to the largest (positive) eigenvalue of S, hence the vector of weights w satisfies the equation $Sw = \lambda_{\max} w$. The existence of the largest eigenvalue is guaranteed by Perron-Frobenius theorem. Let f_j be criteria, let $v_i(f_j)$ be a weight of an alternative i with regard to a criterion j, and let w_i be weights of criteria with regard to the goal, then the weight of an alternative i with regard to the goal is given as $\sum_j v_i(f_j) \cdot w_j$. For the optimal alternative this values attains its maximum.

However, in many real-world situations criteria or alternatives might be interacting and influencing one another. In analytic network process elements are divided into clusters and these clusters form network structures with dependence and feedback. In the first step of ANP elements (criteria) from one cluster are pair-wise compared with regard to elements (criteria) from other clusters, and a supermatrix W is formed. The supermatrix is a block matrix consisting of matrices W_{ij}, where columns of W_{ij} express the relative importance (priority) of elements from a cluster i to elements from a cluster j (columns of W_{ij} are obtained from pair-wise comparisons in the form of eigenvectors). If $W_{ij} = 0$ then a cluster i has no influence on a cluster j.

Intensity of importance	Definition
1	Equal importance
2	Weak
3	Moderate Importance
4	Moderate plus
5	Strong Importance
6	Strong plus
7	Very strong Importance
8	Very, very strong
9	Extreme importance

Table 1 Saaty's fundamental scale. Source: [3]

An element $w_{mn} \in W$ determines direct influence of an element m on an element n. Indirect influences can be determined by raising the supermatrix to the powers, and the final influence is obtained by the limiting process $\lim_{n\to\infty} W^n = W^\infty$ (the weights are in columns). For the convergence (or at least cyclicity) of a supermatrix the stochasticity of W is required.

A structure of the limiting supermatrix W^∞ depends on a network structure (for details see [3] or [4]). For instance, if a goal influences criteria, and criteria influence alternatives as well as criteria themselves, then the supermatrix W has the following form:

$$W = \begin{array}{c} \\ G \\ C \\ A \end{array} \begin{array}{c} \begin{array}{ccc} G & C & A \end{array} \\ \begin{bmatrix} 0 & 0 & 0 \\ W_{21} & W_{22} & 0 \\ 0 & W_{32} & I \end{bmatrix} \end{array},$$

where W_{22} expresses dependencies among criteria. The priority vector (vector of weights) Q of alternatives is obtained from the limiting supermatrix W^∞ in the block W_{31}^∞, which has the form [3]:

$$Q = W_{32}\left(I - W_{22}\right)^{-1} W_{21}$$

Given priorities Q all alternatives can be sorted from the best to the worst.

3 Fuzzy cognitive maps

Cognitive maps (CP) were introduced in 1976 by Axelrod [1] in the context of social and politic decision making. CP consist of concepts (nodes) describing system's behavior connected by edges (arcs) representing their causal relationship. Edges can be assigned three values: 0 (no relationship between two concepts), –1 or '–' (a negative relationship) and 1 or '+' (a positive relationship). In the graph theory nomenclature CP belong among directed graphs.

Fuzzy cognitive maps (FCM), also called fuzzy decision maps, were proposed by B. Kosko in [2], and they can be considered a combination of the fuzzy set theory and neural networks. As relationships among concepts can be 'fuzzy', FCM allow expressing a degree of an influence of one concept on another. To express the degree of an influence of a concept i on a concept j, the mapping $e_{ij} \to [0,1]$ or $e_{ij} \to [-1,1]$ is used, where the higher absolute values of e_{ij} denote the higher influence (the stronger causal relationship). In this paper we define FCM as follows:

Definition 1. *FCM is a tuple (C, E), where C is a set of concepts with cardinality n and E is a square adjacency-matrix of the order n with elements* $e_{ij} \in [0,1]$*, where* e_{ij} *expresses the strength of an influence of a concept i on a concept j.*

Usually, diagonal elements $e_{ii} = 0$ as it is assumed that concepts cannot influence themselves, the rule which we observe too.

In our approach within AHP/ANP framework we use criteria as concepts and their dependence and feedback is modeled by edges representing their relationships. In the following section we provide description of our method for the derivation of criterias' weights.

4 The hybrid eigenvalue-fuzzy cognitive map method for the derivation of global weights of criteria

To derive weights of criteria with interdependence and feedback we propose HEFCM (hybrid eigenvalue-fuzzy cognitive map) method, which proceeds in the following steps:

1. All criteria are pair-wise compared and a reciprocal matrix S is formed. Weights (also called local weights) w of criteria (without dependence and feedback) are established by the standard eigenvalue method. Also, all alternatives are pair-wise compared with regard to all criteria by Saaty's method and their weights are established (these weights in a matrix format correspond to the matrix W_{32} of the supermatrix W in ANP).
2. Fuzzy cognitive map of dependencies and feedbacks among criteria is established and turned into the adjacency (connection) matrix E with elements $e_{ij} \in [0,1]$.
3. Matrix E is converted into normalized matrix E^* with all column sums equal to 1, which is called column stochastic. Column stochasticity of the matrix E^* is necessary for the convergence in the Step 4.
4. The influence of an element i on the element j is given by multiplication of E^*, that is:

$$\lim_{n\to\infty} (E^*)^n = M \tag{1}$$

The convergence of a limit (1) is ensured by column stochasticity of a matrix E^*. If there is no single limit, but a limit cycle consisting of k matrices E_k^*, then the limiting supermatrix M is given by the Césaro summation:

$$\lim_{n\to\infty} \frac{1}{k} \sum_{i=1}^{k} (E_k^*)^n = M \tag{1'}$$

5. The weights are aggregated by the following formula:

$$v = f(w, m) \tag{2}$$

where m is a column vector of M, and $f(w,m)$ is an aggregation operator. Here, we use the arithmetic mean: $v_i = \frac{1}{2}\sum_{i=1}^{n}(w_i + m_i)$. This approach is equivalent to the aggregation formula $v = w + E^* \cdot w$ proposed in [6].

6. The priority vector (global weights) Q of all alternatives with regard to the goal is obtained by the formula:

$$Q = W_{32} \cdot v^T \tag{3}$$

One important advantage of the proposed method is that it does not require threshold functions (logistic, hyperbolic-tangent, sigmoid, etc., see [2] or [6]). In the FCM approach the matrix E is multiplicated by state vectors C_n to (often) generate vector elements larger than 1 or lower than 0, which must be 'cut off' to 1 or 0, respectively, by threshold functions. As a consequence, the result is strongly dependent on a particular threshold function used, see e.g. [6]. In our approach the problem with threshold functions is avoided by the column stochasticity of the matrix E^*. It is worth noting that the very same condition is applied in ANP for the supermatrix W, as it is crucial for the existence of the limiting supermatrix W^∞.

5 The numerical example

In this section we provide the illustration of HEFCM method and its comparison with ANP. In our example we are going to select the best car from two alternatives A and B, when the following interacting criteria for the selection are considered: price (P), safety (S), design (D), equipment (E), and fuel consumption (F). Dependencies among criteria are shown in Figure 4 b).

5.1 HEFCM solution

Step 1: We compare all criteria pair-wise to obtain weights of criteria without dependence and feedback. Results of the comparison are presented in Figure 1. From Figure 1 we derive (normalized) weights of criteria (in the order: P, S, D, E, F) by the eigenvalue method ($\lambda_{\max} = 5.1182$, I.C. = 0.03): w = (0.417, 0.282, 0.153, 0.098, 0.050). Then we compare pair-wise all alternatives with regard to all criteria, see Figure 2 a), and obtain weights of criteria (see Figure 2 b)).

Step 2: We determine the fuzzy cognitive map (the network structure) of criteria, see Figure 4 b), and adjacency matrix E, shown in Figure 3 a).

Step 3: Matrix E is normalized into a matrix E^*, see Figure 3 b).

Step 4: We find the limiting matrix M using (1), the result is presented in Figure 4 a). Weights of criteria emerging from their dependencies are in each column of M, in our case m = (0.273, 0.207, 0.190, 0.293, 0.037).

Step 5: Finally, we aggregate both weights (vectors) w and m via (2): v = (0.345, 0.245, 0.171, 0.195, 0.044).

By comparison of weights w and v it can be seen that the order of criterias' weights changed a little after interdependencies were taken into account as the criterion E (equipment) is now more important than the criterion D (design) due to its higher influence on other criteria. The order of other criteria remained unchanged.

Step 6: Using relation (3) we get Q = (0.569, 0.431), so the optimal selection is the car A.

$$S = \begin{array}{c} \\ P \\ S \\ D \\ E \\ F \end{array} \begin{array}{c} \begin{array}{ccccc} P & S & D & E & F \end{array} \\ \begin{bmatrix} 1 & 2 & 3 & 4 & 6 \\ 1/2 & 1 & 2 & 4 & 5 \\ 1/3 & 1/2 & 1 & 2 & 3 \\ 1/4 & 1/4 & 1/2 & 1 & 3 \\ 1/6 & 1/5 & 1/3 & 1/3 & 1 \end{bmatrix} \end{array}$$

Figure 1 The reciprocal matrix S of pair-wise comparisons of criteria importance

	P	S	D	E	F
A	3	2	1/4	1	1/2
B	1/3	1/2	4	1	2

a)

	P	S	D	E	F
A	0.75	0.67	0.2	0.5	0.33
B	0.25	0.33	0.8	0.5	0.67

b)

Figure 2 a) Pair-wise comparisons of both alternatives with regard to all criteria, b) Saaty's weights of alternatives in the form of a matrix W_{32}

$$E = \begin{array}{c} \\ P \\ S \\ D \\ E \\ F \end{array} \begin{array}{c} \begin{array}{ccccc} P & S & D & E & F \end{array} \\ \begin{bmatrix} 0 & 0.2 & 0.4 & 0.3 & 0 \\ 0.5 & 0 & 0.2 & 0.2 & 0 \\ 0.6 & 0.3 & 0 & 0.1 & 0.2 \\ 0.8 & 0.8 & 0.2 & 0 & 0.2 \\ 0.3 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

a)

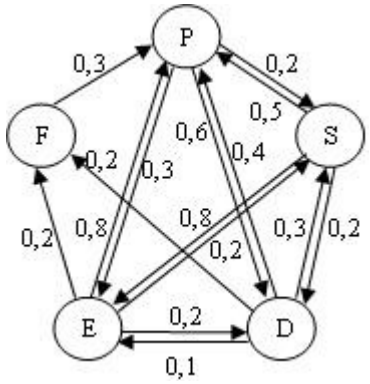
$$E^* = \begin{array}{c} \\ P \\ S \\ D \\ E \\ F \end{array} \begin{array}{c} \begin{array}{ccccc} P & S & D & E & F \end{array} \\ \begin{bmatrix} 0 & 0.154 & 0.5 & 0.5 & 0 \\ 0.227 & 0 & 0.25 & 0.333 & 0 \\ 0.273 & 0.231 & 0 & 0.167 & 0.5 \\ 0.364 & 0.615 & 0.25 & 0 & 0.5 \\ 0.136 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

b)

Figure 3 a) The adjacency matrix *E*, b) the normalized adjacency matrix *E**

$$M = \begin{array}{c} \\ P \\ S \\ D \\ E \\ F \end{array} \begin{array}{c} \begin{array}{ccccc} P & S & D & E & F \end{array} \\ \begin{bmatrix} 0.273 & 0.273 & 0.273 & 0.273 & 0.273 \\ 0.207 & 0.207 & 0.207 & 0.207 & 0.207 \\ 0.190 & 0.190 & 0.190 & 0.190 & 0.190 \\ 0.293 & 0.293 & 0.293 & 0.293 & 0.293 \\ 0.037 & 0.037 & 0.037 & 0.037 & 0.037 \end{bmatrix} \end{array}$$

a)



b)

Figure 4 a) The limiting matrix *M*, b) the fuzzy cognitive map of dependence and feedback among criteria

5.2 ANP solution

The supermatrix *W* of the problem is shown in Figure 6 a), only bold blocks W_{12}, W_{22}, W_{32} and W_{33} are non-zero (see Figure 6 b)). The block W_{22} expressing the relative importance of all criteria with regard to a given criterion was obtained by Saaty's method, pair-wise comparisons are shown in Figure 5. These comparisons were based on the intensity of influence among criteria shown in the fuzzy cognitive map in Figure 4 b).

Because *W* is not stochastic (it has two blocks in one column), blocks W_{22} and W_{32} were compared for their relative importance with the result 'equal importance' (1 on Saaty's scale); hence the stochasticity of *W* was achieved by averaging in columns of blocks W_{22} and W_{32}. New stochastic supermatrix *W´* was then raised to powers and the limiting supermatrix W'^{∞} was found (see Figure 7). Weights (priorities) of alternatives are in the block W'^{∞}_{31} : Q = (0.557, 0.443); hence the alternative A is better than B.

A comparison with the solution of HEFCM method, where Q = (0.569, 0.431), indicates there are only minor differences in alternatives' priorities, and the alternative A is evaluated better by both methods. For completeness, the solution without dependence and feedback (purely AHP) can be obtained from *W* as follows [3]: $Q = W_{32} \cdot W_{21} = (0.598, 0.402)$, which also means that A is better than B.

D	*E*	*F*	*P*	*S*
E	1	1/2	1/4	1/3
F	2	1	1/3	1/2
P	4	3	1	2
S	3	2	1/2	1

E	*D*	*F*	*P*	*S*
D	1	1	1/4	1
F	1	1	1/4	1
P	4	4	1	4
S	1	1	1/4	1

P	*D*	*E*	*S*
D	1	2	3
E	1/2	1	2
S	1/3	1/2	1

S	*D*	*E*	*P*
D	1	1/4	1/3
E	4	1	2
P	3	1/2	1

Figure 5 Pair-wise comparisons of criteria with regard to a given criterion

	G	P	S	D	E	F	A	B
G	W_{11}		W_{12}				W_{13}	
P								
S								
D	W_{21}		W_{22}				W_{23}	
E								
F								
A	W_{31}		W_{32}				I	
B								

a)

	G	P	S	D	E	F	A	B
G	0	0	0	0	0	0	0	0
P	0.417	0	0.320	0.467	0.571	1	0	0
S	0.282	0.163	0	0.277	0.143	0	0	0
D	0.153	0.540	0.122	0	0.143	0	0	0
E	0.098	0.297	0.559	0.095	0	0	0	0
F	0.050	0	0	0.160	0.143	0	0	0
A	0	0.75	0.667	0.2	0.5	0.333	1	0
B	0	0.25	0.333	0.8	0.5	0.667	0	0

b)

Figure 6 a) The general form of the supermatrix for dependence among criteria, b) the supermatrix of the numerical example

	G	P	S	D	E	F	A	B
G	0	0	0	0	0	0	0	0
P	0.417	0	0.160	0.234	0.286	0.5	0	0
S	0.282	0.082	0	0.139	0.072	0	0	0
D	0.153	0.270	0.061	0	0.072	0	0	0
E	0.098	0.149	0.280	0.048	0	0	0	0
F	0.050	0	0	0.080	0.072	0	0	0
A	0	0.375	0.333	0.388	0.25	0.166	1	0
B	0	0.125	0.166	0.612	0.25	0.334	0	0

a)

	G	P	S	D	E	F	A	B
G	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0
F	0	0	0	0	0	0	0	0
A	0.557	0.607	0.601	0.388	0.528	0.470	1	0
B	0.443	0.393	0.398	0.612	0.472	0.530	0	0

b)

Figure 7 a) The stochastic matrix W', b) the limiting supermatrix W'^{∞}

6 Conclusions

In this article we proposed a new hybrid eigenvalue-fuzzy cognitive map method for modeling dependence and feedback among criteria in the ANP framework based on fuzzy cognitive maps. Our method is straightforward, computationally simple and more natural than classic ANP approach. Also, another advantage is that the method does not require the use of ad-hoc threshold functions in the process of adjacency matrix multiplication. The method was demonstrated by the example of the evaluation of criteria weights (for a purchase of a car), but it can be easily extended into cases with dependence among alternatives or among alternatives and criteria. Future research may focus on examination of similarity between HEFCM and ANP in some special cases or in general.

Acknowledgement

The paper was supported by the Grant Agency of the Czech Republic (no. 402090405).

References:

[1] Axelrod, R.: *Structure of Decision: The Cognitive Maps of Political Elites*. Princeton University Press, Princeton, 1976.

[2] Kosko, B.: Fuzzy cognitive maps. *International Journal of Man-Machine Studies* **24** (1986), 65-75.

[3] Ramík, J., Perzina, R.: *Modern methods of evaluation and decision making* (*Moderní metody hodnocení a rozhodování)*. OPF SU, Karviná, 2008.

[4] Saaty, T. L.: *Decision Making with Dependence and Feedback: the Analytic Network Process*. RWS Publications, Pittsburgh, 1996.

[5] Saaty, T. L.: The Analytic Hierarchy and Analytic Network Measurement Processes: Applications to Decisions under Risk. *European Journal of Pure and Applied Math.* **1** (2008), 122-196.

[6] Yu, R., Tzeng, G. H.: A soft computing method for multi-criteria decision making with dependence and feedback. *Applied Mathematics and Computation* **180** (2006), 63-75.

Evolutionary Local Search Algorithm to Solve the Multi-Compartment Vehicle Routing Problem with Time Windows

Jan Melechovský[1]

Abstract. Vehicle routing problems (VRPs) represent an important research stream in applied combinatorial optimization. As the problems emerging from the practice become more constrained and large, more sophisticated approaches are required to address them. This paper tackles the Multiple Compartment Vehicle Routing Problem with Time Windows. This extension of the classical VRP consists of an unlimited homogeneous fleet of vehicles, each being equipped with multiple compartments. Such a configuration enables to load distinct commodities. A customer is characterized by a nonnegative demand of each commodity and a time window allowing its delivery. The goal is to service all customers under the vehicle capacity constraints and time windows constraints with minimal total cost. A metaheuristic solution approach based on evolutionary local search is presented. The performance of the algorithm is evaluated on benchmarks available in the literature.

Keywords: multiple compartment vehicle routing, time windows, local search

JEL classification: C61
AMS classification: 90B06

1 Introduction

Vehicle Routing Problem (VRP) is a combinatorial optimization problem modeling situations in which a nonnegative demand of customers must be satisfied by a fleet of vehicles available at a central depot. It is assumed that each customer can be serviced by exactly one vehicle and the total load of an vehicle cannot exceed its capacity. Each vehicle route must start and end at the depot. The objective is to minimize the total travel cost or alternatively to minimize the number of vehicles used in the solution. The problem is known to be strongly NP-hard. One of many natural extensions of VRP's formulations is the VRP with Time Windows. In this problem, a constraint stating that each customer must be visited within a given time interval must be additionally taken into account. See for example [9] for a survey of VRP and VRPTW variants and solution approaches.

The Multi-Compartment VRP and VRPTW (MCVRP and MCVRPTW respectively) represent generalizations of the classical problem. In MCVRPTW each customer requests a delivery of a set of different products and these products must be transported separately. Vehicles have multiple compartments with limited capacity. Each compartment is dedicated to contain one commodity. The constraint that the demand of each customer must be totally satisfied with a single visit of a vehicle is somewhat relaxed in MCVRPTW. This constraint is reduced only to each commodity while two different commodities are allowed to be delivered in two vehicles. Thus multiple visits of a customer are possible.

Although the problem arises quite often in the practice, it has been seldom studied in the literature. A typical application of MCVRP is the delivery of petroleum products to petrol stations using tank trucks with two compartments (see [2, 3]). In [1] the authors propose a branch-and-price algorithm to solve the MCVRP in which vehicles are equipped with different tanks. Another practical application is the delivery of groceries to convenience stores [4]. Each product requires different temperature, for example low temperature compartment and normal compartment. Animal food distribution to farms is an example of MCVRP mentioned in [5]. The authors propose a memetic algorithm and a tabu search

[1]University of Economics, Prague, Department of Econometrics, nám. W. Churchilla 4, 130 67, Praha 3 , Czech Republic, e-mail: jan.melechovsky@vse.cz

to address the problem. A collection of different kinds of waste is another application studied in [7]. The solution approach is based on a guided local search.

The principal contribution of the paper is the proposition of a metaheuristic algorithm to solve a difficult vehicle routing problem. Moreover, to the best of our knowledge, the problem (MCVRPTW) has not been studied in the literature.

2 Problem formulation

The problem is defined on an undirected graph $G(V, E)$ with $V = \{0, 1, \ldots, n\}$ denoting the set of customer nodes (plus the depot node with index 0) and E representing the set of edges. Each edge $e(i, j)$ has a cost c_{ij} and a travel time t_{ij}. It is assumed that both sets of edge values satisfy the triangle inequality. The depot can supply m different products which can be delivered to customers using a fleet K of identical vehicles with m compartments. Each compartment $p \in \{1, 2, \ldots, m\}$ has a capacity Q_p. Each customer node i has a demand d_{ip} of product p, possibly equal to 0 if product p is not ordered by customer i. A time window $[a_i, b_i]$ is associated with each customer specifying the time slot in which the customer can be visited. The time windows are considered as hard, i.e. earlier arrival than a_i results in a waiting time and later arrival than b_i is not allowed. Moreover, a service time s_i is associated with each visit of customer i. A feasible solution of the problem consists of a set of routes, each starting and ending at the depot, such that the total demand of each product and customer is fully satisfied, the capacity of each vehicle compartment is respected, each delivery to a customer takes place within the given time window and each product is delivered in a single route to each customer. The objective is to find such a feasible solution that the total travel costs are minimized.

3 Evolutionary local search algorithm

As already mentioned the problem is NP-hard and exact solution approaches are rather limited to problems of small size. We are not aware of any application of exact methods to MCVRPTW, but since the problem is harder to solve than VRPTW and the best exact algorithms can deal with instances containing 100 customers and the computing times might grow up to several hours, we can conclude that metaheuristics could be more effective in solving problems of practical size.

The proposed algorithm is an evolutionary local search algorithm (ELS). This method was originally proposed by Merz and Wolf in [6] for a peer-to-peer problem in telecommunications. It extends the classical iterated local search (ILS) which, starting from an initial solution, successively generates child solutions using a perturbation mechanism and local search. The perturbation mechanism serves as a diversification tool while the local search intensifies the search in the current solution neighborhood. ELS additionally generates multiple child solutions and only the best child is kept. The perturbation modifies the solution in a random manner with a parameter r denoting the number of changes made in the solution. If the best child solution improves the best solution found so far the parameter is set to its minimal value r_{min}. If not, r is augmented by a predefined step (e.g. $r = \min(r + 1, r_{max})$).

ELS performs the following steps:

1. Initilize solution x with a simple constructive heuristic.
2. Set $r = r_{min}$ and the best solution $X_{best} = X$.
3. Repeat for a given number of max iterations:
   (a) Initialize best child solution: set $Cost(X'_{best}) = \infty$.
   (b) Repeat for a given number of child solutions:
      i. Initialize child solution: $X' = X_{best}$.
      ii. Perturb X' using perturbation parameter r.
      iii. Apply local search procedure to X'.
      iv. Update best child solution: set $X'_{best} = X'$ if $Cost(X') < Cost(X'_{best})$.
   (c) Update best found solution and perturbation parameter:

- if $Cost(X'_{best}) < Cost(X_{best})$:
  - Set $X_{best} = X'_{best}$.
  - Set $r = r_{min}$.
- else: $r = \min(r + 1, r_{max})$

First, an initial solution is determined by a simple heuristic. The perturbation parameter r is initialized to its minimal value and the initial solution enters the ELS loop. ELS performs a given number of iterations with the same value of r. In every iteration, a given number of child solutions is generated using the perturbation mechanism and local search. Only the best child solution is kept in the memory. If the best child solution is better than the best solution found so far, the latter is updated and r is reset to r_{min}. If not, r is incremented by a predefined step.

Initial heuristic

The initial solution is obtained with a simple construction heuristic. The solution is initialized with an empty route. Then all customers with some unsatisfied demand are scanned and the best insertion of a customer into a route is determined. If the customer cannot be inserted into any already existing route, the insertion into a new route is considered instead. The criterion for the insertion is the least increase of travel costs. The procedure terminates when the demand of all customers is fully satisfied.

Perturbation

The perturbation mechanism is one of the two routines of ELS that modify the incumbent solution. It plays the role of a diversification tool in the general ELS framework since it performs several random operations on a solution. Thus it can be interpreted as a mutation operator used in genetic algorithms. The procedure is controlled with a parameter r denoting the number of operations to be performed. In our implementation the operation is a removal and relocation of a customer. For a given route of a solution, r customer nodes are randomly selected and removed from the route. Each of the r removed customers is then tested for a feasible insertion into some of the remaining routes of the solution. If such feasible insertion is detected, the customer is relocated to its new position. If not, new route visiting only this single customer is created.

Local search

Local search is applied to the solution modified by the perturbation mechanism. Its purpose is to improve the solution using a set of operators. It intensifies the search and the improved solution represents a local optimum within the given solution neighborhood. Together with the perturbation mechanism it enables the algorithm to explore effectively the solution space and find potentially good solutions.

Each local search operator defines a solution neighborhood, i.e. the set of solutions that can be obtained by applying the operator. Hence the definition of an operator has a crucial effect to the quality of found solutions. Usually one local search operator is not sufficient to explore the solution space effectively. Combining more operators can increase the probability of finding high quality solutions. On the other hand, the computational complexity shall be also considered.

The implemented local search procedure relies on three operators:

- Path Exchange – exchanges two sub-paths between two routes,
- Relocate – relocates one node within the same route or between two routes,
- Swap – swaps a pair of nodes between two routes.

The operators are applied sequentially with the first improvement strategy – i.e. the first detected improving move is performed. The search stops when no improving move can be found by any operator. Each move must be checked for the feasibility. This involves the time windows as well as the capacity. The time windows feasibility can be checked in $O(1)$ if an information of the maximum feasible shift of

each node visit is kept in the memory. However, the update of this information must be done for each node and requires $O(n)$. The capacity must be checked for all vehicle compartments if the move involves two routes. It can be done in $O(1)$ except for Path Exchange where it needs $O(n)$ steps.

4 Preliminary results

The algorithm was coded in C++ and the computational experiments were carried out on a PC equipped with 2.9 GHz dual core processor and 2 GB of RAM. To the best of our knowledge there are no benchmarks proposed in the literature. Therefore the testing environment was derived from existing VRPTW instances.

Test instances

The test instances were derived from the famous VRPTW instances proposed by Solomon in [8]. The data set contains 56 problems divided into six sets: C1, R1, RC1, C2, R2 and RC2. Each instance contains 100 customers plus the depot. The nodes are distributed in a 100×100 square around the depot which is positioned in the middle. Customers are clustered in sets C1 and C2, randomly distributed in sets R1 and R2 and mixed clustered and randomly distributed in sets RC1 and RC2. The position of nodes is identical for all instances within the same set. Only the time window differs. Instances in sets C2, R2 and RC2 are characterized by wider time windows and are reputed as harder to solve.

In [5] the authors have proposed a manner how to derive a MCVRP instances from VRP data sets. We have adapted their idea for the MCVRPTW case. The number of compartments was set to 2. We have considered only sets C1, R1 and RC1 in the preliminary study. The first data set was obtained by splitting the customer demand and the capacity of compartments into two equal parts. The advantage is that any solution feasible for VRPTW remains feasible for MCVRPTW.

The second data set was designed in a way ensuring asymmetric demands of each product. The demand of the first product is calculated as $d_{i1} = d_i/k$, where d_i is the demand of customer i in the original instance and k is a random number from $[3, 5]$. The demand of the second product is $d_{i2} = d_i - d_{i1}$. The values are rounded to nearest integer. The capacity of each compartment is calculated as a function of the total demand and the average demand of each product.

Parameters setting

The ELS algorithm requires only few parameters to be set. The total number of ELS iterations was set proportionally to the number of customers: $N_{max} = 2n$ and the number of child solutions generated within each ELS iteration was set to 10. The perturbation parameter r ranged from 1 to 4.

Results

The computational experiments are still in process so we cannot provide the reader with a detailed analysis of the efficiency of the algorithm. We have performed several tests on the first set of instances, for which the optimal VRPTW solutions are known. The average gap between the ELS solution and the optimum was 4.1 %, 5.8 % and 5.4 % for data sets C1, R1 and RC1 respectively. The average computational times were 79.5 s, 106.7 s and 126.7 s. The average gap is relatively close to the optimum but still it is remarkable that the proposed algorithm is not designed to solve pure VRPTW. We beleive that with further improvements of the local search the solutions might be closer to optima. The average computational times seem to be reasonable when taking into account the difficulty of the problem.

5 Conclusions

The paper introduced a generalization of VRPTW in which multiple vehicle compartments and requested products are taken into account. The problem has not been studied in the available literature. An algorithm based on the evolutionary local search was proposed to address the problem. First results give

an optimistic expectations regarding the overall performance of the solution approach. However, further experiments must be carried out in order to obtain better analysis of the performance. Next steps in the research will be focused on the implementation of other operators in the local search and eventually on the design of another metaheuristic framework. A post-optimization phase might be also implemented in order to better explore the promising areas of the solution space.

Acknowledgements

The research was fully supported by the Czech Science Foundation, project reg. number P402/12/P635.

References

[1] Avella, P., Boccia, M., and Sforza, A.: Solving a fuel delivery problem by heuristic and exact approaches, *European Journal of Operational Research* **152** (2004), 170 – 179.

[2] Brown, G. G., Ellis, C. J., Graves, G. W., and Ronen, D.: Real-time, wide area dispatch of mobil tank trucks, *Interfaces* **17** (1987), 107 – 120.

[3] Brown, G. G., and Graves, G. W.: Real-time dispatch of petroleum tank trucks, *Management Science* **27** (1981), 19 – 32.

[4] Chajakis, E. D., and Guignard, M.: Scheduling deliveries in vehicles with multiple compartments, *Journal of Global Optimization* **26** (2003), 43 – 78.

[5] El Fallahi, A., Prins, C., and Wolfler Calvo, R.: A memetic algorithm and a tabu search for the multi-compartment vehicle routing problem, *Computers & Operations Research* **35** (2008), 1725 – 1741.

[6] Merz, P., and Wolf, S.: Evolutionary local search for the super-peer selection problem and the p-hub median problem. In: *Lecture notes in computer science* (T. Bartz-Beielstein et al., eds.), volume 4771, Springer Berlin/Heidelberg, 2007, 1 – 15.

[7] Muyldermans, L., and Pang, G.: On the benefits of co-collection: Experiments with a multi-compartment vehicle routing algorithm, *European Journal of Operational Research* **206** (2010), 93 – 103.

[8] Solomon, M. M.: *Vehicle routing and scheduling with time window constraints: models and algorithms.* PhD thesis, University of Pennsylvania, 1983.

[9] Toth, P., and Vigo, D., eds., *The vehicle routing problem.* Society for Industrial and Applied Mathematics, Philadelphia, 2001.

Optimal allocation of government debt for the Czech Republic: Managing vulnerability of debt service charges to macroeconomic shocks

Aleš Melecký[1], Martin Melecký[2]

Abstract: Although debt sustainability issues are often analyzed in respect of the level of government debt, unexpected increases in debt services charges can substantially change the dynamics and accumulation of government debt. These unexpected increases could be a result of exposures to exchange rate, interest rate and refinancing risks that the government has assumed when allocating its debt across a spectrum of available financial (debt) instruments. Therefore public debt managers aim to minimize the vulnerability of public debt to macroeconomic shocks by carefully allocating debt through new issuance and buybacks across currencies, interest rates, maturities and markets, among others. In this paper, we build on the methodology proposed by Giavazzi and Missale (2004) to estimate the optimal debt allocation for the Czech Republic taking into account the structure of the Czech economy and the macroeconomic shocks that it is faced with. We develop policy recommendation to improve on the existing central government debt allocation and general government asset and liability management based on our estimations results.

Keywords: government debt allocation, risk management, macroeconomic shocks, Czech Republic.

JEL Classification: H63, E44
AMS Classification: 49K30

1. Introduction

The concept of fiscal sustainability is most often analyzed in the context of sustainable fiscal policies and the needed budgetary adjustments. However unsustainable fiscal stance can arise also as a result of risky government (public) debt allocations or misalignments of government assets and liabilities. Or alternatively, a better, more cost-efficient debt allocation can be implemented to support fiscal consolidation efforts of a government while taking on the same amount of risk.

IMF and World Bank (2001) include a comprehensive review of public debt management policies and approaches for policy makers, including formulation and properties of public debt management objectives, the underlying institutional framework and possible coordination issues, ensuing formulation of the debt management strategy, attributes of a sound risk management framework, and other important areas of public debt management. Melecky (2012a) presents and empirical analysis of possible drives behind different formulations of public debt management strategies across a sample of countries. From a practical risk management perspective, Buera and Nicolini (2002) find that the size of financial transactions the government must undertake each period to replicate state contingent bonds is very large and increases dramatically with number of states. Further, Melecky (2012b) provides a review of policy approaches to choosing the currency structure of foreign-currency debt in view of the fact that historically the exchange rate risk is the most important risk for the debt managers in emerging market economies. Gerard and Gilson (2001) show in a simple two country model how an exchange rate regime can

[1] Technical University of Ostrava, Czech Republic, Department of Economics, e-mail: ales.melecky@vsb.cz.
[2] Technical University of Ostrava, Czech Republic, Department of Economics, e-mail: m.melecky@gmail.com.

influence public debt structure. Melecky (2010) then develops empirical framework that debt managers could use when deciding on the currency allocation of public external debt building on a set of synchronization indicators of exchange rate volatility. Further, Concerning the Czech Republic, Matalik and Slavik (2005) describe historical development and the current state of public debt management in the Czech Republic. They conclude that public debt management should be included as part of the state treasury management. Overall there is a lack of empirical studies dealing with public debt management issues in the Czech Republic.

This paper uses the approached proposed by Giavazzi and Missale (2004) to estimate the optimal allocation of the Czech government's debt in the view of the government's fiscal consolidation plan and its probability of successful implementation. The considered instruments for allocation of the government debt are the domestic floating debt, domestic fixed rate debt, domestic inflation-indexed debt, and euro-denominated foreign debt. This government debt management problem has not been research much in the existing empirical literature concerning Central European economies, and the Czech Republic in particular. This paper thus attempts to fill in this gap.

We find that in view of the openness and structure of the Czech economy and the macroeconomic shocks it is facing, the Government should allocate most of its debt, about 57 percent, into domestic fixed rate bonds. The second biggest allocation, of about 40 percent, should be made towards foreign euro-denominated bonds. Further, about 3 percent of the debt should be optimally allocated into domestic floating rate notes, while zero allocation is suggested for inflation-indexed bonds. The comparison of our estimated optimal government debt allocation with the actual one followed by the Czech Ministry of Finance shows important differences. In sum, these differences may imply that the Ministry of Finance could, at the same level of volatility of debt service charges, lower its borrowing cost by making some changes in its debt allocation. Our preliminary results suggest that this would be best done, over a medium term, by increasing the allocations into foreign euro-denominated bonds while decreasing the allocation into domestic fixed-rate bonds.

The remainder of the paper is organized as follows. Section two discusses the adopted modeling approach. Section three describes the employed data series. Section four explains the calibration of key parameters for estimation of the optimal public debt allocation according to Giavazzi and Missale (2004). Section five discusses obtained results in the context of the actual government debt allocation applied by the Czech Ministry of Finance. Section six concludes.

2. The modeling approach

When analyzing the optimal allocation of government debt for the Czech Republic, we follow the approach of Giavazzi and Missale (2004). In view of the objective function for fiscal policy and debt management, including the size of expected fiscal consolidation $E_t\left(A_{t+1}-\Delta B^T_{t+1}\right)$ and the probability of its failure $\Pr$, Giavazzi and Missale work out the optimal solution for the allocation of government debt across short-term floating-rate debt, s^*, foreign-currency denominated debt, q^*, inflation-indexed debt, h^*, and long-term fixed-rate debt, $(1-s^*-q^*-h^*)$:

$$s^* = \frac{(\eta_y + B_t)}{B_t}\frac{Cov(y_{t+1}i_{t+1})}{Var(i_{t+1})} + \frac{(\eta_\pi + B_t)}{B_t}\frac{Cov(\pi_{t+1}i_{t+1})}{Var(i_{t+1})}$$
$$-q^*\frac{Cov(e_{t+1}i_{t+1})}{Var(i_{t+1})} - h^*\frac{Cov(\pi_{t+1}i_{t+1})}{Var(i_{t+1})} + TP_t\frac{\sqrt{2\Pr}}{1-\sqrt{2\Pr}}\frac{E_t\left(A_{t+1}-\Delta B^T_{t+1}\right)}{B_t Var(i_{t+1})} \quad (1)$$

$$q^* = \frac{(\eta_y + B_t)}{B_t}\frac{Cov(y_{t+1}e_{t+1})}{Var(e_{t+1})} + \frac{(\eta_\pi + B_t)}{B_t}\frac{Cov(\pi_{t+1}e_{t+1})}{Var(e_{t+1})}$$
$$-s^*\frac{Cov(e_{t+1}i_{t+1})}{Var(e_{t+1})} - h^*\frac{Cov(\pi_{t+1}e_{t+1})}{Var(e_{t+1})} + FP_t\frac{\sqrt{2\Pr}}{1-\sqrt{2\Pr}}\frac{E_t\left(A_{t+1} - \Delta B_{t+1}^T\right)}{B_t Var(e_{t+1})} \quad (2)$$

$$h^* = \frac{(\eta_y + B_t)}{B_t}\frac{Cov(y_{t+1}\pi_{t+1})}{Var(\pi_{t+1})} + \frac{(\eta_\pi + B_t)}{B_t}$$
$$-q^*\frac{Cov(e_{t+1}\pi_{t+1})}{Var(\pi_{t+1})} - s^*\frac{Cov(\pi_{t+1}i_{t+1})}{Var(\pi_{t+1})} + IP_t\frac{\sqrt{2\Pr}}{1-\sqrt{2\Pr}}\frac{E_t\left(A_{t+1} - \Delta B_{t+1}^T\right)}{B_t Var(\pi_{t+1})} \quad (3)$$

where η_y and η_π are elasticity of government budget to GDP with respect to output and inflation, and B_t is government debt to GDP. *Cov*(.) stands for covariance and *Var*(.) for variance of corresponding variables. *Pr* denotes probability that the adopted stabilization plan fails, and E_t (A_t-ΔB_t^T) is the planned reduction in debt-to-GDP ratio over period T. TP_t, FP_t, IP_t represents the term premium, the foreign exchange premium on Czech koruna vis-à-vis the euro, and the inflation premium respectively.

3. Data description

In order to estimate the optimal government debt allocation for the Czech Republic according to (1)-(3), we first focus on computing the unknown covariances and variances of real GDP, inflation, the interest rate and the nominal exchange rate. We use quarterly data series for the Czech Republic from the first quarter of 1996 to the fourth quarter of 2011. The data are obtained from the Czech National Bank (CNB), Eurostat, and the Czech Statistical Bureau (CSU). The computation of the individual data series is described in more detail in Table 1 including the data sources for the individual variable.

Variable	Notes	Source
Real GDP growth	Annualized percentage change in real GDP	ARAD – CNB
Inflation	Annualized percentage change in harmonized CPI (2005=100)	EUROSTAT
Interest rate	3Month PRIBOR	ARAD – CNB
Nominal exchange rate	Log of nominal exchange rate CZK/EUR	EUROSTAT
Government Budget	total government revenues-total government expenditures	CSU

Table 1 Data description

Note: Data from 1996Q1 to 2011Q4.

4. Calibration of parameters

In the next step, we calibrate the remaining parameters of equations (1)-(3). The subject parameters, their description and numerical calibrations are presented in Table 2:

Parameter	Description	Value
η_y	Elasticity of gov. budget to GDP with respect to output	0.04
η_π	Elasticity of gov. budget to GDP with respect to inflation	-0.07
B_t	Government debt to GDP by end-2011	37.00
TP_t	Term premium	1.54
FP_t	Foreign Exchange Premium on CZK	-3.95
IP_t	Inflation premium	0.10
Pr	Probability that stabilization plan fails	2.00
$E_t(A_t\text{-delthaB}_t^T)$	Planned reduction in debt-to-GDP ratio over T	1.00

Table 2 Calibration of parameters in equations (1)-(3)

Source: Authors' calculations in percent

The semi-elasticities of the government budget to GDP ratio with respect to output and inflation were estimated as the respective correlations over 1996Q1-2011Q4. Note that the elasticity to inflation is negative which is somewhat puzzling and deserves further investigation which we leave for further research.[3] The government debt to GDP ratio, *Bt*, was set to 37 percent, the level at the end of 2011. The term premium was calculated as the average of the difference between the yield of 10-year government bond and the yield of 6-month money market rate (assumed to be equivalent to the 6-month Treasury bill rate). The foreign exchange premium on the Czech koruna vis-à-vis the euro is computed as the average of the difference between the percentage change in the CZK/EUR nominal exchange rate and the interest rate differential between the 3-month PRIBOR rate and 3-month EURIBOR rate. The PRIBOR rate was obtained from the CNB database and the EURIBOR from Eurostat.

The inflation premium is calculated as the average of the difference between actual CPI inflation at time *t* and the expected CPI inflation conditional on an information set dated *t*-1. The AR(1) process was used to generate expected inflation for simplicity. This is because data on inflation expectations are not readily available or inflation linked bonds traded. We leave more sophisticated treatment of expected inflation for further research. Note that another simple approximation of inflation expectations could be achieved by using the CNB inflation target at a given time, assuming perfect credibility of CNB's monetary policy and its inflation target. Alternatively, fast learning of the economic agents would need to be in place to ensure this approximation holds during a monetary policy-driven disinflationary period, as experienced by the Czech Republic.

The probability that a given stabilization (fiscal consolidation) plan may fail was set at 2 percent following Giavazzi and Missale (2004). In further research, we will consider more thoroughly the track record of the Czech government in adhering to its announced stabilization plans, most notably those involving significant consolidation.

The consolidation plan itself involving a planned reduction in debt-to-GDP ratio, $E_t\left(A_{t+1} - \Delta B_{t+1}^T\right)$, was set following Giavazzi and Missale to 1 percent annual reduction in government debt to GDP. Note that the Czech Republic is currently envisaging continuing, though declining, fiscal deficits and debt accumulation with the balance budget planned to be reached in 2015.

[3] This finding is robust to using annualized q-to-q inflation, y-to-y inflation or detrended inflation. The negative correlation of budget balance to GDP with inflation prevails.

5. Discussion of results

Employing the calibration discussed above and using the *fsolve* Matlab function to solve the equations (1)-(3) of three unknowns, we arrive at the estimation results that are presented in Table 3:

Considered Allocation	Estimated Optimal	Actual, December 2011
s^*	3%	10%
q^*	40%	18%
h^*	0%	0%
$(1-s^*-q^*-h^*)$	57%	72%

Table 3 Debt allocation of government debt for the Czech Republic

Source: Authors' calculations; MoF development of the government debt.

Our estimates suggest that in view of the historical macroeconomic developments, the structure of the Czech economy and the macroeconomic shocks it is facing, the Government should allocate most of its debt, about 57 percent, into domestic fixed rate bonds. This is to support its objective of fiscal consolidation and achievement of targeted outcomes while bearing in mind possible failures in exact implementation of the planned consolidation. The second biggest allocation, of about 40 percent, should be made towards foreign euro-denominated bonds. About 3 percent of the debt should be allocated into domestic floating rate notes, while zero allocation should be made towards inflation-indexed bonds.

The last column of Table 3 shows the actual allocation of Czech government's debt (in securities) across the considered categories by end-2011. It is interesting to observe that much more weight, 72 percent, is given to the allocation into domestic fixed-rate bonds which suggest more conservative (risk-averse) debt management preferences and possible additional objective of domestic debt market development (IMF and WB, 2001). Similarly, less weight is given by the Ministry of Finance to allocation into foreign currency bonds, about 18 percent, in comparison to the 40 percent allocation into euro-denominated bonds suggested by our estimates.

Although the presented research should be further developed and the robustness of the acquired results properly assessed, the comparison of our estimated optimal government debt allocation and the actual one followed by the Czech Ministry of Finance shows important differences. In sum, these differences could imply that the Ministry of Finance could achieve better outcomes in terms of possibly lower cost at the same level of risk by making some changes in its debt allocation. Our very preliminary results suggest that this would be best done, over a medium term, by increasing the allocations into the euro-denominated bonds on account of a decreased allocation into domestic fixed-rate bonds.

6. Conclusion

This paper acknowledges that unexpected increases in debt services charges due to risky allocation of government debt can substantially change the dynamics and accumulation of government debt. The risky allocation can derive from an excessive exposure of the government to exchange rate, interest rate and refinancing risks. This paper thus

aimed to exploit the approach of Giavazzi and Missale (2004) in order to provide some insight and guidance to public debt managers in the Czech Republic to minimize the vulnerability of government debt to macroeconomic shocks while achieving adequate efficiency in terms of borrowing costs. Based on our preliminary estimates, we found that the Czech government should allocate most of its debt, about 57 percent, into domestic fixed rate bonds -- as it currently does but to a much larger extent. Further, it should allocate about 40 percent of its debt into foreign euro-denominated bonds and about 3 percent of its debt into domestic floating rate notes. Zero allocation is suggested for inflation-indexed bonds. Our preliminary results imply that the Ministry of Finance could lower its borrowing cost, at the best possible cost-risk trade-off, by increasing its debt allocations into euro-denominated bonds while decreasing its debt allocation into domestic fixed-rate bonds, over the medium term.

Acknowledgements

Financial support from the Czech Science Foundation GA402/10/1046 is gratefully acknowledged.

References

[1] Buera, F. and Nicolini, J. P.: Optimal Maturity of Government Debt without State Contingent Bonds. Universidad Torcuato Di Tella, Department of Economics Working Papers 016, 2002.
[2] Gerard, M. and Gilson, N.: Public debt structure and exchange rate regime. Dipartimento di Economia Politica dell'Università di Milano Bicocc Working Paper, 2001.
[3] Giavazzi, F. and Missale, A.: Public Debt Management in Brazil. National Bureau of Economic Research, Inc., Cambridge, NBER Working Paper No.10394, 2004.
[4] IMF and the World Bank: Guidelines for Public Debt Management.Washington D.C., 2001.
[5] Matalik, I. and Slavik, M.: Debt Management in the Czech Republic (formation in the 1990s and the current state). *Prague Economic Papers* **1** (2005), 33-50.
[6] Melecky, M.: Formulation of Public Debt Management Strategies: An Empirical Study of Possible Drivers. *Economic Systems* **36** (2012), 218-234.
[7] Melecky, M.: Choosing The Currency Structure Of Foreign-Currency Debt: A Review Of Policy Approaches. *Journal of International Development* **24** (2012), 133-151.
[8] Melecky, M.: Currency Allocation of Public External Debt and Synchronization Indicators of Exchange Rate Volatility. *Comparative Economic Studies* **52** (2010), 104-129.

National Efficiency Evaluation of Visegrad Countries in Comparison with Austria and Germany by Selected DEA Models

Lukáš Melecký[1], Michaela Staníčková[2]

Abstract. The paper deals with an application of Data Envelopment Analysis (DEA) method to multi-criteria efficiency evaluation of the Visegrad Four countries (V4) in comparison with selected advanced European Union's (EU) countries – Austria and Germany. The aim of the paper is to analyse a degree of efficiency achieved in individual countries which is perceived as a reflection of the level of competitive potential in the reference years 2000 and 2010. The theoretical part of the paper is devoted to the fundamental bases of competitiveness and the methodology of factor analysis and DEA method. The empirical part is aimed at measuring the effectiveness of selected countries by selected DEA models. When applying factor analysis and DEA method, we used indicators, which are part of the Country Competitiveness Index (CCI) created by World Economic Forum (WEF) and EU. Indicators included in CCI are interrelated, therefore we use factor analysis for assessment of internal relations between indicators and for reduction of their high number to a smaller number of variables, but at a minimum loss of information contained in the original variables. Factor analysis allows to use a structure of common factors of all variables and create factors including the most important and convenient indicators for national efficiency evaluation. Results of factor analysis are used for calculations of selected DEA models – basic CCR and BCC models and additive SDM, FDH and FRH models. The DEA method evaluates the efficiency how countries are able to transform their inputs into outputs. Therefore, efficiency of countries can be considered as a 'mirror' of national competitiveness. The final part of the paper offers a comprehensive comparison of results obtained by using of all calculated DEA models.

Keywords: Competitiveness, DEA method, efficiency, factor analysis, model BCC/CCR/FDH/FRH/SBM, Visegrad Four countries.

JEL Classification: C61, C67, O11, P51, Y10
AMS Classification: 62H20, 62H25, 90C05, 93B15, 93D25

1 Introduction

European Union is a heterogeneous unit with significant economic and social disparities between its Member States and their regions. The support of cohesion and balanced regional development together with increasing level of EU competitiveness belong to the temporary *key development objectives* of the EU. In relation to *competitiveness, performance* and *efficiency* are *complementary objectives*, which determine the long-term development of states and regions. The *measurement, analysis and evaluation of productivity changes, efficiency and level of competitiveness are controversial topics* and have enjoyed great interest among researchers.

1.1 Concept of Competitiveness

The *definition of competitiveness* faces to the *lack of mainstream view* for understanding this term. Competitiveness remains a concept that can be understood in different ways and levels despite widespread acceptance of its importance. Although there is *no uniform definition and understanding* of this concept, competitiveness remains one of the *fundamental criteria of economic performance evaluation* and it is also seen as a reflection of area (country/region) success in a wider (international/interregional) comparison. The concept of competitiveness is distinguished at different levels - *microeconomic, macroeconomic and regional*. Anyway, there are some differences between these three approaches; see e.g. [8]. In original meaning the concept of competitiveness was applied only to companies and corporate strategies. *Competitiveness of companies* is understood as the ability to

[1] VŠB-Technical University of Ostrava, Faculty of Economics, Department of European Integration, Sokolská třída 33, 701 21 Ostrava, lukas.melecky@vsb.cz.
[2] VŠB – Technical University Ostrava, Faculty of Economics, Department of European Integration, Sokolská třída 33, 701 21 Ostrava, michaela.stanickova@vsb.cz.

provide products and services as well as or more effectively than their main competitors; see e.g. [9]. There is not such a standardized definition and understanding of *national competitiveness* in comparison with the competitiveness at the microeconomic level. One of the most common interpretations understood national competitiveness as the ability to produce goods and services in the country that are able to successfully face international competition, and people can enjoy a growing and sustainable living standards [7]. Macroeconomic concept of national competitiveness cannot be fully applied at the regional level because the regional competitiveness is much worse and less clear defined; between these two concepts is a big difference; see e.g. [8]. *Regional competitiveness* can be understood as the result of joint efforts on the most productive use of internal resources development in the interaction with the use of external resources and development opportunities focused on sustainable increases in production potential.

1.2 Evaluation of Competitiveness

Competitiveness is most commonly evaluated by *decomposition of aggregate macroeconomic indicators.* To compare a level of national competitiveness we can use the databases performed by *Institute for Management Development* (IMD) – the *World Competitiveness Yearbook* (WCY), and *World Economic Forum* (WEF) – the *Global Competitiveness Report* (GCR). Decomposition of aggregate macroeconomic indicators of international organizations is the most commonly used approach mainly at regional level, as well as *comprehensive analysis* aimed at identifying the *key factors* of regional development, productivity and economic growth; see e.g. [11]. *EU competitiveness* can be measured also by indicators of *EU' growth strategies* (Lisbon strategy – Structural indicators, Strategy Europe 2020 – Indicators of Europe 2020) or by *macro-econometric modelling* with creation of an econometric panel data model; see e.g. [7]. There is continuity between approach of EU and WEF in *EU Country/Regional Competitiveness Index*; see e.g. [1]. Another approach is evaluation by *DEA method*, which measures national/regional efficiency and subsequent national/regional competitive potential; see e.g. [10].

2 Multivariate Methods of Competitive Potential Measurement

The most common quantitative methods convenient for a high number of multivariate measured variables can be identified as *multivariate statistical methods*. Multivariate analysis is an ever-expanding set of techniques for data analysis that encompasses a wide range of possible research situation [6]. Between collection of multivariate statistical methods we can include e.g. Factor analysis, Cluster analyses or DEA method.

2.1 Theoretical Basis of Factor Analysis

Factor analysis is a collection of methods used to examine how underlying constructs influence the responses on a number of measured variables. Factor analysis is a method for investigating whether a number of variables of interest $Y_1, Y_2,\ldots Y_n$, are *linearly related* to a smaller number of unobservable factors $F_1, F_2,\ldots F_k$. If we suggest that one measured variable Y_1, is function of two underlying factors, F_1 and F_2, then it is assumed that Y variable is linearly related to the two factors F, as follows in equation (1):

$$Y_1 = \beta_{10} + \beta_{11}F_1 + \beta_{12}F_2 + e_1. \tag{1}$$

The error terms e_1, serves to indicate that the hypothesized relationships are not exact. In the special vocabulary of factor analysis, the parameters $\beta_{i,j}$ are referred to as *loadings.* For example, β_{12} is called the loading of variable Y_1 on factor F_2. There is generally a wide range of literature based on factor analysis. For example, a hands-on how-to approach can be found in Stevens [11]; more detailed technical descriptions are provided in Hair and Black [6]. De Coster [4] posted, that there are basically *two types of factor analysis*: exploratory and confirmatory. *Exploratory factor analysis* (EFA), which is applied in this paper, attempts to discover the nature of the constructs influencing a set of responses. *Confirmatory factor analysis* (CFA) tests whether a specified set of constructs is influencing responses in a predicted way.

The main *applications* of factor analytic techniques are (1) to *reduce* the number of variables and (2) to *detect structure* in the relationships between variables that is to *classify variables*. Therefore, factor analysis is applied as a *data reduction* or *structure detection method*. Factor analyses are performed by examining the pattern of *correlations* between the observed measures. Measures that are highly correlated (either positively or negatively) are likely influenced by the same factors, while those that are relatively uncorrelated are likely influenced by different factors. The primary objectives of an *EFA* are to determine (1) The number of common factors influencing a set of measures and (2) The strength of the relationship between each factor and each observed measure. There are *seven usual basic steps* to performing *EFA*, used in the empirical analysis of the paper: (1) Collection of measurement variables; (2) Obtain the correlation matrix between each of variables; (3) Selection of the number of factors for inclusion; (4) Extraction of initial set of factors; (5) Rotation of factors to a final solution; (6) Interpretation of factor structure; (7) Construction of factor scores for further analysis.

2.2 Fundamental Background of DEA Method

The performance analysis provided by *Data Envelopment Analysis* (DEA) method can be used for evaluating territorial development efficiency with respect to the territorial factor endowment. DEA was first proposed by Charnes, Cooper and Rhodes [2] in 1978. Since that time, researchers in a number of fields have quickly recognized that it is an excellent and easily used methodology for modelling operational processes for performance evaluations. DEA is based on *Farrel model* for measuring the effectiveness of units with one input and one output, which expanded *Charnes, Cooper, Rhodes* (CCR model) and *Banker, Charnes, Cooper* (BCC model), in advanced *Slack-Based Model* (SBM), *Free Disposal Hull* (FDH) and *Free Replicability Hull* (FRH) models [3].

DEA is gaining importance as a tool for evaluating and improving the performance of a set of peer entities called *Decision Making Units* (DMUs)) which convert multiple inputs into multiple outputs. DEA is a *multi-criteria* productivity analysis model for measuring relative efficiency and providing comparison of a homogenous set of DMUs. The DMUs are usually characterized by several inputs that are utilized for producing several outputs, but their performances are different. DMU is *efficient* if the observed data correspond to testing whether the DMU is on the imaginary '*production possibility frontier*'. All other DMU are simply *inefficient*. The best-practice units are used as a reference for evaluation of other group units. The *aim of DEA method* is to examine DMU if they are *effective* or *not effective* by size and quantity of consumed resources by produced output [3].

3 Application of Multivariate Methods to Efficiency Analysis

3.1 Efficiency Analysis Background

Based on the facts above, it is possible to determine the initial *hypothesis of the analysis*. The hypothesis is based on the assumption that *countries achieving best results in efficiency are countries best at converting inputs into outputs (best using of competitive advantages) and therefore having the greatest performance and productive potential*. DEA is applied to 4 countries within the V4 Group– Czech Republic (CZ), Hungary (HU), Poland (PL) and Slovakia (SK), and to 2 selected advanced EU countries – Austria (AT) and Germany (DE). The efficiency analysis starts from building database of measurable indicators that are part of a common approach of WEF and EU in the form of *Country Competitiveness Index* (CCI). The aim of this approach is to develop a rigorous method to benchmark national competitiveness and to identify key factors which drive competitiveness performance of countries. The reference to CCI is the well-established *Global Competitiveness Index* (GCI) by WEF. Eleven pillars of GCI are used for CCI constructing and may be grouped according to the different dimensions (*input versus output aspects*) of national competitiveness they describe. *Methodology of CCI is thus suitable for national competitiveness evaluation by Factor analysis and DEA method* [1]. The *68 indicators* selected for the CCI framework are all of quantitative type (hard data) and consist of several database sources. In this paper, database analysis consists of *66 selected indicators* – 38 of them are inputs and 28 outputs. The reference period is set across years 2000 and 2010. We do not use all indicators included in CCI because not all indicators were available for the whole period for each explored country, but for some indicators we have found comparable indicators. The pillars and used indicators are listed in Annex – Table 1.

For calculations of economic efficiency of evaluated countries, we have used *10 selected DEA models* with multiple inputs and outputs: *1. CCR input oriented model assuming constant returns to scale (CRS), 2. CCR output oriented model assuming CRS, 3. BCC input oriented model assuming variable returns to scale (VRS), 4.BCC output oriented model assuming VRS, 5. SBM additive model not-focusing on input and output assuming CRS, 6. SBM additive model not-focusing on input and output assuming VRS, 7. FDH input oriented model, 8.FDH output oriented model, 9. FRH input oriented model, 10. FRH output oriented model.* For solution of DEA models, we have used software tools based on solving linear programming problems, e.g. Solver in MS Excel, such as the *DEA Frontier*. Assuming *6 countries*, each with *m* inputs and *r* outputs, efficiency of a test country *q* is obtained by solving equations (2) [5]. Given the extensive equations of each model, only basic CCR model is shown for illustrative purposes, for *CCR input oriented model* with CRS the following equation (2) [5]:

$$\max\ z = \sum_{i}^{r} u_i y_{iq}, \qquad (2)$$

on conditions:

$$\sum_{i}^{r} u_i y_{ik} \le \sum_{j}^{m} v_j x_{jk}, k = 1,2,\ldots,n,$$

$$\sum_{j}^{m} v_j x_{jq} = 1,$$

$$u_i \ge \varepsilon,\ i = 1,2,\ldots,r,$$

$$v_j \ge \varepsilon,\ j = 1,2,\ldots,m.$$

Where:

z	the coefficient of efficiency of unit U_q;
ε	infinitesimal constant;
v_j	weights assigned to j-th input;
u_i	weights assigned to the i-th output;
x_{jk}	value of j-th input of unit U_k;
x_{jq}	value of j-th input of unit U_q;
y_{ik}	value of i-th output of unit U_k;
y_{iq}	value of i-th output of unit U_q;
m	inputs;
r	outputs.

Basic DEA models, *primary CCR input/output oriented models*, assume CRS. *BCC input/output oriented models* consider VRS (decreasing, increasing or constant). VRS enable better identify more efficient units, because VRS provides a more realistic expression of economic reality and factual relations and activities existing in countries. CCR and BCC models evaluate efficiency of countries for any number of inputs and outputs. The *coefficient of efficiency* (CE) is ratio between the weighted sum of outputs and the weighted sum of inputs. Each country selects input and output weights that maximize its *efficiency score*. The CE takes values in interval <0,1>. In *DEA models aimed at inputs* the CE of efficient countries equals 1, while the CE of inefficient countries is less than 1. In *DEA models aimed at outputs* the CE of efficient countries equals 1, but the CE of inefficient countries is greater than 1. In formulation of *SBM additive models is not necessary to distinguish between a focus on inputs and outputs*. In SBM models, the CE of efficient units equals 0, because it is the sum of additional variables for inputs and outputs (s^+ and s^-), which express the distance from the efficient frontier. The sum of additional variables for inputs and outputs is lower, evaluated countries is closer to the efficient frontier package and thus has a higher degree of efficiency, and otherwise [3]. The basic idea of *FDH model* is unconvexity of set of production possibilities. This means that evaluated unit can be only relatively compared towards really existing units. For comparison with CCR and BCC models, it should be added that limits of efficiency rate is similar to these models, and it depends on model orientation on inputs or outputs. Rate of efficiency, obtained by FDH models, is generally higher than in CCR and BCC models. This is due to the possibility that a production unit is dominated not only by specific production units of set of units (in the case of CCR and BCC models), as well as convex combinations of these units. A simple extension of FDH model is *FRH model*, which unlike FDH model, allows evaluated unit compares with multiplied combinations of other units [5].

3.2 Competitiveness Factors Measurement by Factor Analysis

For utilization of above mentioned sources, set of *66 variables* was compiled. In order to ensure comparability between different countries, all variables have to be relativized, and these variables thus entered into analysis. In process of data preprocessing is necessary to make their *standardization* (normalization), thus to unify their standards. The most commonly used method of standardization is to transform data into *Z-scores*. Based on used data standardization method, Pearson's correlation coefficient was chosen as a measure of correlation. The ideal would be case in which correlation degree of variables do not fall below 0.3. Like would not fall below 0.3, correlation coefficients should appropriate variables or vice versa exceed 0.9. On basis of defined conditions, *15 variables* for inputs and *13 variables* for outputs were excluded. Relevant new database consists now of *38 indicators – 23 input* and *15 output indicators*, illustrated in Table 1 in Annex also with excluded variables in crossed font.

After a relatively complex process of variables selection, the core of factor analysis follows. Statistical package SPSS (in our case *IBM SPSS Statistics – Version 20*) provides a wide range of methods for factors extraction. In this paper we have chosen specifically modified *method of principal components* because of higher number of variables. By its application to input set of variables, an estimate of *factor/component matrix* (often called also as matrix of factor loads) has been provided. Number of factors has been predefined in input parameters by determining the value of own number to a value greater than 1.0. Own number of a particular factor indicates the amount of total variability explained by just this factor. Very frequently criterion for finding the *optimal number of factors*, the percentage of total variance explained collectively by selected factors, is used. For an imaginary boundary of quality solution is widely accepted 70 % of explained variability. In our case, *five dominating factors for inputs* explained *100 %* of total variability in years 2000 and 2010, which can be considered as very satisfactory result. In the case of *outputs – four dominating factors* explained *95.168 %* of total variability in year 2000 and *94.188 %* of total variability in year 2010, which can be considered also as very satisfactory results.

The *optimal number of factors* is already known (*5 factors for inputs* in years 2000 and 2010, and *4 factors for outputs* in years 2000 and 2010), their interpretation still proceed not. One of yet unnamed conditions is that each factor has influence the most of variables, while each of variables, if it is possible, and should depend on the fewest number of factors. Further step is to rotate of factors or factorial axes, which task is just to maximize the load of each variable in one of the extracted factors, while her loads under other factors are substantially minimized. The *Varimax method* of rotation, which rotates the coordinate axes in the direction of maximum variance, has been used. Results clearly show that target of rotation was almost completely fulfilled. Only a few variables are now characterized by high loads in more than one factor and total structure of factor matrix is considerably simplified. For interpretation, those variables were identified as relevant, factor loadings exceeded the 0.4. This frontier was marked as convenient by *Stevens* [11]. Based on results of correlation and factor analysis, we could proceed to DEA method. Indicators for inputs and outputs, depending on their level of significance for competitiveness of evaluated countries, were divided by results of factor analysis in 2000 and 2010.

3.3 Evaluation of National Efficiency by DEA Models

The initial hypothesis of efficiency being a mirror of competitive potential *was confirmed* through analysis as illustrated in following Tables 1 and 2 in years 2000 and 2010. In the case of national efficiency evaluation was found out that in used DEA models were comparable results in all V4 countries, but also in Austria and Germany. *At national level, it is evident that levels of efficiency of individual V4 countries are lower in CCR models than in BCC, FDH and FRH models* (except Austria and Germany, which were evaluated to be efficient in all models during referred period). *This fact confirms theory that in BCC models with VRS, the CE reach higher values and higher number of evaluated countries is classified as efficient. This has been also confirmed in SBM models with VRS by higher number of evaluated countries identified as highly efficient compared to SBM models with CRS. This fact is also confirmed in FDH and FRH models*, because these models relatively compare inputs and outputs of evaluated countries towards really existing countries, and not to virtual country.

The overall evaluation of efficiency of V4 countries, Austria and Germany shows that the best results achieved 2 of 6 countries in years 2000 and 2010. The best results are predictably achieved by economically powerful countries which were *efficient in all used DEA models* during the whole referred period. Therefore, according to hypothesis, these countries should have the *greatest competitive potential*. Efficient countries – *Austria* and *Germany*, are highlighted by dark grey colour in Tables. The efficient countries are followed by a group of countries which are *slightly inefficient*. These countries do not achieved efficiency equal to 1 in *CCR, BCC, FDH* and *FRH models* or low sum of values of additional variables in *SBM* models, but their efficiency indices reached consistently highly effective values close during referred period (coloured by light grey colour in Tables). These countries are *Czech Republic*, *Slovakia* and *Poland in all used DEA models*. Only Hungary *was classified as inefficient in all used DEA models*, so it shows low competitive potential and development perspective (coloured by ultra-light grey colour and italics in Tables 1 and 2).

Tables also show position of individual V4 countries and Austria and Germany within selected DEA models in terms of the order of achieved values of efficiency coefficients in CCR, BCC, FDH and FRH models or sum of values of additional variables in SBM models in years 2000 and 2010. The overall evaluation of individual countries shows that best results, in terms of efficiency in all used DEA models, *Austria* and *Germany* have reached and are ranked in *first place*. These countries thus effectively utilize their competitive advantages. In *second place*, there is *Czech Republic*, which was evaluated as slightly inefficient with high level of competitive potential. *Slovakia* and *Poland* are ranked in *third* and *fourth place* because they have reached the lower values of the CE in CCR, BCC, FDH and FRH models, and higher sum of values of additional variables in SBM models. *Hungary* was ranked in *last – fifth place*, because it was classified as *inefficient* with the lowest values of the CE in CCR, BCC FDH and FRH models, and the highest sum of values of additional variables in SBM models.

Country 2000	CCR IO	CCR OO	BCC IO	BCC OO	SBM CRS	SBM VRS	FDH IO	FDH OO	FRH IO	FRH OO	Absolute Rank of Country
AT	1	1	1	1	0	0	1	1	1	1	**1.**
DE	1	1	1	1	0	0	1	1	1	1	**1.**
CZ	0,969	1,089	0,975	1,056	3 750	850	0,986	1,032	0,991	1,026	**2.**
HU	*0,901*	*1,123*	*0,915*	*1,109*	*1 456 003*	*404 589*	*0,926*	*1,087*	*0,945*	*1,071*	**5.**
PL	0,934	1,107	0,942	1,089	65 893	35 025	0,961	1,072	0,972	1,054	**4.**
SK	0,944	1,097	0,958	1,072	51 236	28 567	0,978	1,059	0,983	1,048	**3.**

Note: * IO = input oriented model, OO = output oriented model
Table 1 Results of Selected DEA Models in Year 2000 According to Coefficients of Efficiency
Source: Own calculation and elaboration, 2012

Country 2010	CCR IO	CCR OO	BCC IO	BCC OO	SBM CRS	SBM VRS	FDH IO	FDH OO	FRH IO	FRH OO	Absolute Rank of Country
AT	1	1	1	1	0	0	1	1	1	1	**1.**
DE	1	1	1	1	0	0	1	1	1	1	**1.**
CZ	0,985	1,041	0,993	1,029	1 126	252	0,995	1,015	0,998	1,004	**2.**
HU	*0,927*	*1,101*	*0,937*	*1,095*	*901 969*	*226 946*	*0,949*	*1,071*	*0,967*	*1,062*	**5.**
PL	0,952	1,089	0,961	1,078	51 256	19 925	0,978	1,062	0,985	1,041	**4.**
SK	0,966	1,079	0,975	1,059	37 595	17 261	0,986	1,041	0,991	1,028	**3.**

Note: * IO = input oriented model, OO = output oriented model
Table 2 Results of Selected DEA Models in Year 2010 According to Coefficients of Efficiency
Source: Own calculation and elaboration, 2012

4 Conclusion

Based on DEA analysis has been found out that in evaluated countries is a *distinct gap* between economic and social standards, so *differences still remain*. Development in V4 countries has a trend towards advanced countries, such as Austria and Germany. There was a growth in their performance, increasing trend in effective use of their advantages and improve in competitive position. But most countries experienced also a decline in their performance (outputs decline as a result of declines in inputs) as a result of economic crisis. The recent economic crisis has seriously threatened the achievement of sustainable development in the field of competitiveness. The crisis has underscored importance of competitiveness – supporting economic environment to enable national economies to better absorb shocks and ensure solid economic performance going into the future.

Acknowledgements

This paper was created under SGS project (SP2012/153) of Faculty of Economics, VŠB-TU of Ostrava.

References

[1] Annoni, P., and Kozovska, K.: *EU Regional Competitiveness Index 2010*, Publication Office of the European Union, 2010.
[2] Charnes, A., Cooper, W., Rhodes, E.: Measuring the efficiency of decision making units. *European Journal of Operational Research* **2** (1978), 429-444.
[3] Cook, W. D., and Zhu, J.: *Data Envelopment Analysis: Modelling Operational Processes and Measuring productivity*. Kluwer Academic Publishers, Boston, 2008.
[4] DeCoster, J.: *Overview of Factor Analysis* [online]. 1998. <http://www.stat-help.com/notes.html>.
[5] Fiala, P.: *Operační výzkum: Nové trendy*. Professional Publishing, Praha, 2010.
[6] Hair, J. F., and Black, W.C. et al.: *Multivariate Data Analysis*. Prentice Hall, 2009.
[7] Hančlová, J. et al.: *Makroekonometrické modelování české ekonomiky a vybraných ekonomik EU*. VŠB-Technical University, Ostrava, 2010.
[8] Krugman, P.: Competitiveness: A Dangerous Obsession. *Foreign Affairs* **2** (1994), 28-44.
[9] Porter, M. E.: The Economic Performance of Regions. *Regional Studies* **6**/**7** (2003), 549-578.
[10] Staníčková, M., and Skokan, K.: Evaluation of the EU Member States Efficiency by Data Envelopment Analysis Method. *International Journal of Mathematical Models and Methods in Applied Sciences* **2** (2012), 349-357.
[11] Stevens, J.: *Applied multivariate statistics for the social sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1986.

Annex

Dimension	Pillar	Indicator*
Inputs	1. Institution	In: **Political Stability** Out: ~~Voice and Accountability, Government Effectiveness, Regulatory Quality, Rule of Law, Control of Corruption~~
	2. Macroeconomic Stability	In: **Harmonized Index of Consumer Prices, Gross Fixed Capital Formation** Out: ~~Income, Saving and Net Lending/Net Borrowing, General Government Gross Debt, Total Intramural Research & Development Expenditure, Labour Productivity per Person Employed~~
	3. Infrastructure	In: **Railway transport - Length of Tracks, Air Transport of Passengers, Volume of Passenger Transport, Volume of Freight Transport** Out: ~~Motorway Transport - Length of Motorways, Air Transport of Freight~~
	4. Health	In: **Healthy Life Expectancy, Infant Mortality Rate, Cancer Disease Death Rate, Heart Disease Death Rate, Suicide Death Rate** Out: ~~Hospital Beds, Road Fatalities~~
	5. + 6. Primary, Secondary and Tertiary Education, Training and Lifelong Learning	In: **Mathematics-Science-Technology Enrolments and Graduates, Pupils to Teachers Ratio, Financial Aid to Students, Total Public Expenditure at Primary Level of Education, Total Public Expenditure at Secondary Level of Education, Total Public Expenditure at Tertiary Level of Education, Participants in Early Education, Participation in Higher Education, Early Leavers from Education and Training, Accessibility to Universities** Out: ~~Lifelong Learning~~
	9. Indicators for Technological Readiness	In: **Level of Internet Access** Out: ~~E-government Availability~~
Outputs	7. Labour Market Efficiency	In: **Labour productivity, Male employment, Female employment, Male unemployment, Female unemployment, Public expenditure on Labour Market Policies** Out: ~~Employment rate, Long-term unemployment, Unemployment rate~~
	8. Market Size	In: **Gross Domestic Product** Out: ~~Compensation of employees, Disposable income~~
	10. Business Sophistication	In: **Gross Value Added in sophisticated sectors, Venture capital (expansion- replacement)** Out: ~~Employment in sophisticated sectors, Venture capital (investments early stage)~~
	11. Innovation	In: **Human resources in Science and Technology, Total patent applications, Employment in technology and knowledge-intensive sectors, Employment in technology and knowledge-intensive sectors-by gender, Employment in technology and knowledge-intensive sectors-by type of occupation, Employment in technology and knowledge-intensive sectors-by level of education** Out: ~~Human resources in Science and Technology – Core, Patent applications to the EPO, Total intramural R&D expenditure, High-tech patent applications to the EPO, ICT patent applications to the EPO, Biotechnology patent applications to the EPO~~

Table 1 Indicators of Inputs/Outputs in Period 2000-2005-2010 Relevant to Factor Analysis
Note: * Number of indicators was decreased after correlation in inputs from 38 to 23, in outputs from 28 to 15
Source: [1]; own calculation and elaboration, 2012

Shapley value of simple cooperative games with fuzzy coalitions applied on the real voting data

Elena Mielcová[1]

Abstract. The main aim of this article is to compare the results of classical Shapley value concept with results of the Shapley value extended to the cooperative games with fuzzy coalitions applied on the real data of the cooperative simple game – in this case the data from the voting in the Lower House of the Czech Parliament 2002-2012. One of the most intriguing tasks is to describe real system problems using mathematical tools. As the real systems are full of an uncertainty, there are several tools incorporating the uncertainty into the classical models, for example the theory of probability, the theory of fuzzy sets, or the theory of rough sets. In this article, the probabilistic approach to the Shapley value was chosen. Results indicate the improvement in predictability of the real power of existing political coalitions comparing to the values based on the classical Shapley value.
.

Keywords: Shapley value, power distribution, coalitions, Czech Parliament.

JEL Classification: C71
AMS Classification: 91B12

1 Introduction

The history of the classical approach to the cooperative games evaluation goes to early 1950s, when Lloyd S. Shapley published his famous article "*A value for n-Person Games"* [4]. He presented his concept on the transferable utility cooperative games in characteristic-form function. He defined three axioms (symmetry, efficiency, and additivity) which such a value in any game should fulfill. He had shown that there exists one unique value fulfilling all three axioms, now called Shapley value.

The application of the Shapley value to the evaluation of coalitions in simple games with non-transferable utility function, and with coalitional structure – simply called voting games – was done in 1954 when L. S. Shapley and M. Shubik published their article "*A method for evaluating the distribution of power in a committee system*" [5]. They showed principle of the a-priori evaluation of power distribution in a simple game committee system.

Since then, the applications Shapley value were done in different fields, for example in economy, political science, medicine or computer science. The Shapley value was recalculated and new equivalent function form notations of the Shapley value were postulated.

Lately, as the concept of uncertainty was introduced to the field of game theory, the Shapley value was recalculated for different levels of uncertainty. For example Mareš [2] in his book "Fuzzy cooperative games" derived the Shapley value for cooperative games with fuzzy pay-offs. Butnariu and Kroupa [1] narrowed their approach to the n-person games with fuzzy coalitions under the condition of utility aggregation, Yu and Zhang [9] derived the Shapley value for fuzzy bi-cooperative games.

The main aim of this article is to compare the results of classical Shapley value concept with results of the Shapley value extended to the cooperative games with fuzzy coalitions applied on the real data of the cooperative simple game – in this case the data from the voting in the Lower House of the Czech Parliament 2002-2012. In this article, I have chosen the probabilistic approach to the Shapley value of the simple voting game.

This article is organized as follows: in the next part I will describe original idea and the basic definition of the Shapley value, as well as the basic definitions of fuzzy coalitions and the application of the classical Shapley value concept on the real data. The results of calculations and comparison of the a-priori Shapley-Shubik power index and the Shapley value of fuzzy coalitions with real voting outcomes are given in the third section. The final discussion and the list of references end the paper.

[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, the Czech Republic, mielcova@opf.slu.cz..

2 Basic definitions

In the original Shapley article [4], the Shapley value is defined on the class of games in characteristic function form expressed for each subset S of universe of players U. The characteristic function v is expected to be superadditive ($v(S) \geq v(T) + v(S-T)$ for all $T, S \subseteq U$), $v(\varnothing) = 0$. Shapley defined a carrier of v to be any set $N \subseteq U$ with $v(S) = v(N \cap S)$, and $\Pi(U)$ be the set of permutations of U. The three Shapley axioms, "symmetry" (for all $\pi \in \Pi(U), \phi_{\pi}(\pi v) = \phi_i(v)$), "efficiency" (for each carrier N of v, $\sum_N \phi_i(v) = v(N)$), and "additivity" (for all independent u, v $\phi(v+u) = \phi(v) + \phi(u)$) are sufficient to determine a unique ϕ for each player *i* for all games [4]:

$$\phi_i(v) = \sum_{i \in S \subseteq N} \frac{(s-1)!(n-s)!}{n!} [v(S) - v(S \setminus i)] \qquad (1)$$

where $N \subset U$ is any finite carrier of v. This value ϕ is called Shapley value.

The calculation of the Shapley value on the power evaluation of coalitions in simple games (games in characteristic function form v for which $v(S) \in \{0,1\}$ for all S) with non-transferable utility function is based on the existence of marginal players in coalitions [5]. Moreover, the value of players in such a game is dependent on the creation of a-priori coalitions. The calculation of these values, called Shapley-Shubik power indices together with Shapley-Shubik power indices with a-priori coalitions are described in [5, 6, 7, 8].

Let $N = \{1,2,\ldots,n\}$ denotes set of all players. By a coalition we mean any subset of *N*. Then a fuzzy coalition is a vector $A = (A(1), A(2), \ldots, A(n))$ with coordinates $A(i) \in \langle 0,1 \rangle$ called the membership degree of player *i* in coalition *A* . In the description of a voting game in the parliamentary voting, usually political parties are expected to be players of the voting game. The membership of such a player in coalition thus can be the probability of the political party occurrence in the coalition. Moreover, the characteristic function v can reach a value from the interval $\langle 0,1 \rangle$, the higher the value, the higher is the probability of the coalition to be winning. On such values, the common Shapey value (1) can be applied.

In order to evaluate the party success during a parliamentary period, the index of party success can be constructed by comparing party decision with the outcome of parliamentary voting. The party A success index is defined as the ratio of decisions of the Lower House that were the same as the party A decisions to all decisions during the parliamentary period [3]:

$$I^{A}_{success} = \frac{\text{number of party A decisions identical with parliamentary decisions}}{\text{number of parliamentary decisions}} \qquad (2)$$

The party decision is derived from the votes of party members using simple majority rule; the coefficient of party success is a value from the interval $\langle 0,1 \rangle$; the higher the coefficient, the higher ratio of party decisions was the same as the whole voting body decision.

3 Shapley value of real voting data

The analysis is based on the roll-call voting data from the Lower House of the Czech Parliament 2002-2012, that means two complete electoral period (2002/2006, and 2006-2010), and part of the latest electoral period (2010-2012). Data are available at the official Parliamentary site www.psp.cz. Votes of legislators are collected into voting vectors; one voting vector contains voting outcomes for one bill of all 200 members. The outcome of every vote for every member can be one element of the set {A, N, 0, Z, M}, which indicates member status or preferences: N – "no", A – "yes", Z – "present, abstain", 0 –"absent", M –"absent, excused". Every bill to be passed needs at least as many "yes" votes as the quota. The quota is based on the sum of all present legislators, which means on the sum of all legislators with the voting outcome from the set {N, A, Z}. Hence the outcome "present, abstain" in this analysis is reclassified to "no" outcome [3]. Data for the 2002-2006, 2006-2010, and 2010-2012 Lower House parliamentary periods cover 4797, 8470, and 3286 voting vectors, respectively. The last session – session 36 from the data set was on 3/14/2012. During the studied period, there were eight political parties active in the Czech Parliament; their names followed by the abbreviation used in this article are Civic Democratic Party (ODS), Czech Social Democratic Party (CSSD), Christian and Democratic Union – Czechoslovak People's Party (KDU-CSL), Czech and Moravian Communist Party (KSCM), Freedom Union (US), Green Party (SZ), TOP 09 (TOP09), Věci veřejné (VV).

A-priori Shaply-Shubik power indices

The values of an a-priori Shapley-Shubik power index are dependent on the number of legislators in political parties. The values of the Shapley-Shubik power indices for political parties present in the 2002-2006 Lower House together with Shapley-Shubik index with a-priori coalition (CSSD, KDU-CSL and US), and with the index of success are given in Table 1.The correlation coefficients of the index of success with the calculated Shapley-Shubik power index, and with the Shapley-Shubik power index with a-priori coalitions are -0.073, and 0.664, respectively. Both coefficients are not statistically significant at 0.05 levels.

	CSSD	ODS	KSCM	KDU-CSL	US
Shapley-Shubik power index	**0.400**	**0.233**	**0.233**	**0.067**	**0.067**
Shapley-Shubik power index with a-priori coalition of CSSD, KDU-CSL, and US	**0.749**	**0**	**0**	**0.125**	**0.125**
Index of Success	**0.928**	**0.596**	**0.697**	**0.904**	**0.857**

Table 1 Shapley-Shubik power indices and the index of success for political parties present in the 2002-2006 Lower House of the Czech Parliament. Source: own calculations.

Similarly, the values of the Shapley-Shubik power indices for political parties present in the 2006-2010 Lower House together with the index of success are given in Table 2.The correlation coefficient of the index of success with the calculated Shapley-Shubik power index is 0.704 and it is not statistically significant.

	CSSD	ODS	KSCM	KDU-CSL	SZ
Shapley-Shubik power index	**0.2833**	**0.3667**	**0.2833**	**0.0333**	**0.0333**
Index of Success	**0.8130**	**0.8072**	**0.7222**	**0.7485**	**0.6776**

Table 2 Shapley-Shubik power indices and the index of success for political parties present in the 2002-2006 Lower House of the Czech Parliament. Source: own calculations.

The values of the Shapley-Shubik power indices for political parties present in the 2010-2012 Lower House together with Shapley-Shubik index with a-priori coalition (CSSD, KDU-CSL and US), and with the index of success are given in Table 3.The correlation coefficients of the index of success with the calculated Shapley-Shubik power index, and with the Shapley-Shubik power index with a-priori coalitions are -0.143, and 0.843, respectively. Both correlation coefficients are not statistically significant at 0.05 levels.

	CSSD	ODS	KSCM	TOP09	VV
Shapley-Shubik power index	**0.3**	**0.3**	**0.133**	**0.133**	**0.133**
Shapley-Shubik power index with a-priori coalition of ODS, TOP09, and VV	**0**	**0.530**	**0**	**0.235**	**0.235**
Index of Success	**0.5663**	**0.9714**	**0.5426**	**0.9744**	**0.9680**

Table 3 Shapley-Shubik power indices and the index of success for political parties present in the 2010-2012 Lower House of the Czech Parliament. Source: own calculations.

Shapley value of the 2002-2010 Lower House based on real votes

In general, the Shapley value for this type of the real voting game is not a good predictor of the future success of political parties – players of the voting game. In order to study the possible predictions of future success of political parties, the Shapley values of the Lower House voting games based on (1) using real voting data (2002-2010) were calculated. Theoretical and real weights, as well as theoretical and real characteristic functions for the 2002-2006, and 2006-2010, respectively, are given in Tables 4, and 6. The same values could be calculated for the 2010-2012 parliamentary period however the period is still not over, so the values could not describe the real situation properly.

Tables contain membership degrees of all players in all possible coalitions, as well. These values are substantial for the future success prediction. Membership degrees are calculated as an average value of according votes for every coalition. Political parties are players of this voting game, thus the membership functions of players not present in coalitions might be positive.

The normalized Shapley values of political parties calculated with respect to votes in the 2002-2006 Lower House of the Czech parliament are given in Table 5. The correlation coefficients of the calculated Shapley values with the coefficient of success are 0,39, and 0.722 and are not statistically significant at 5% level.

	Theoretical weight	Theoretical v(S)	Real weight	Real v(S)	A(A)	A(B)	A(C)	A(D)	A(E)
0	0	0	0.038	0	0.141	0.080	0.069	0.072	0.085
A	0.2	0	0.067	0	0.886	0.052	0.108	0.058	0.074
B	0.2	0	0.003	0	0.081	0.820	0.086	0.283	0.267
C	0.2	0	0.073	0	0.088	0.066	0.933	0.043	0.036
D	0.2	0	0.001	0	0.240	0.205	0.146	0.658	0.158
E	0.2	0	0.004	0	0.216	0.152	0.107	0.184	0.696
AB	0.05	1	0.001	0.833	0.774	0.734	0.158	0.168	0.178
AC	0.05	0	0.136	0.158	0.947	0.031	0.895	0.029	0.033
AD	0.05	0	0.003	0	0.892	0.097	0.063	0.720	0.163
AE	0.05	0	0.009	0.045	0.857	0.049	0.089	0.156	0.807
BC	0.05	1	0.014	0.746	0.090	0.841	0.924	0.175	0.101
BD	0.05	0	0.004	0.263	0.069	0.895	0.168	0.883	0.240
BE	0.05	0	0.001	0	0.107	0.883	0.104	0.194	0.853
CD	0.05	0	0.001	0	0.085	0.241	0.810	0.768	0.210
CE	0.05	0	0.002	0.111	0.150	0.173	0.831	0.156	0.717
DE	0.05	0	0.003	0	0.165	0.187	0.052	0.798	0.868
ABC	0.0333	1	0.005	1	0.835	0.813	0.948	0.202	0.132
ABD	0.0333	1	0.004	1	0.850	0.827	0.169	0.844	0.283
ABE	0.0333	1	0.001	1	0.915	0.759	0.190	0.232	0.790
ACD	0.0333	1	0.004	0.950	0.893	0.117	0.896	0.806	0.147
ACE	0.0333	1	0.003	0.833	0.931	0.105	0.832	0.117	0.760
ADE	0.0333	0	0.008	0.175	0.910	0.159	0.076	0.825	0.899
BCD	0.0333	1	0.018	1	0.143	0.951	0.911	0.844	0.193
BCE	0.0333	1	0.005	1	0.108	0.903	0.937	0.280	0.837
BDE	0.0333	1	0.076	0.675	0.053	0.969	0.105	0.971	0.967
CDE	0.0333	0	0.002	0.400	0.226	0.266	0.842	0.832	0.822
ABCD	0.05	1	0.027	1	0.784	0.848	0.893	0.816	0.304
ABCE	0.05	1	0.006	1	0.760	0.795	0.854	0.342	0.842
ABDE	0.05	1	0.047	1	0.912	0.934	0.067	0.957	0.964
ACDE	0.05	1	0.006	1	0.919	0.180	0.914	0.717	0.733
BCDE	0.05	1	0.116	0.996	0.114	0.948	0.892	0.942	0.926
ABCDE	0.2	1	0.313	1	0.859	0.920	0.931	0.928	0.915

Table 4 Theoretical and real weights, theoretical and real characteristic functions, and membership degrees of all players in all possible coalitions for the 2002-2006 Lower House of the Czech Parliament. Codes: ODS=A, CSSD=B, KSCM=C, KDU-CSL=D, US=E. Source: own calculations.

	ODS	**CSSD**	**KSCM**	**KDU-CSL**	**US**
Normalized Shapley value	0.120	0.444	0.217	0.138	0.082
Normalized Shapley value with a-priori coalition of CSSD, KDU-CSL, and US	0.016	0.639	0.029	0.198	0.117

Table 5 Shapley value of political parties present in the 2002-2006 Lower House of the Czech Parliament based on the voting outcomes. Source: own calculations.

	Theoretical weight	Theoretical v(S)	Real weight	Real v(S)	A(A)	A(B)	A(C)	A(D)	A(E)
0	0	0	0.043	0	0.093	0.088	0.079	0.121	0.065
A	0.2	0	0.019	0.018	0.830	0.055	0.070	0.193	0.055
B	0.2	0	0.013	0.072	0.065	0.868	0.131	0.103	0.055
C	0.2	0	0.047	0	0.069	0.148	0.893	0.098	0.043
D	0.2	0	0.005	0	0.202	0.153	0.129	0.648	0.116
E	0.2	0	0.011	0	0.142	0.104	0.103	0.183	0.814
AB	0.05	1	0.009	1	0.807	0.822	0.121	0.265	0.176
AC	0.05	1	0.008	0.647	0.834	0.124	0.851	0.229	0.101
AD	0.05	0	0.020	0.186	0.906	0.069	0.060	0.800	0.137
AE	0.05	0	0.008	0.028	0.892	0.057	0.035	0.223	0.913
BC	0.05	0	0.148	0.409	0.037	0.945	0.963	0.076	0.046
BD	0.05	0	0.004	0.452	0.122	0.894	0.150	0.752	0.058
BE	0.05	0	0.003	0	0.079	0.889	0.139	0.200	0.834
CD	0.05	0	0.003	0	0.156	0.145	0.874	0.668	0.088
CE	0.05	0	0.004	0	0.152	0.132	0.858	0.204	0.883
DE	0.05	0	0.004	0	0.222	0.166	0.131	0.694	0.831
ABC	0.0333	1	0.039	1	0.778	0.834	0.869	0.306	0.169
ABD	0.0333	1	0.018	1	0.848	0.868	0.142	0.796	0.117
ABE	0.0333	1	0.007	0.983	0.844	0.855	0.126	0.332	0.912
ACD	0.0333	1	0.006	0.961	0.921	0.111	0.861	0.800	0.166
ACE	0.0333	1	0.002	0.944	0.878	0.106	0.850	0.248	0.942
ADE	0.0333	0	0.080	0.544	0.963	0.055	0.037	0.924	0.954
BCD	0.0333	1	0.019	0.949	0.108	0.943	0.965	0.777	0.087
BCE	0.0333	1	0.016	0.863	0.094	0.931	0.940	0.200	0.863
BDE	0.0333	0	0.003	0.393	0.166	0.876	0.149	0.771	0.899
CDE	0.0333	0	0.002	0.063	0.193	0.178	0.879	0.735	0.824
ABCD	0.05	1	0.087	1	0.858	0.896	0.897	0.817	0.186
ABCE	0.05	1	0.040	1	0.798	0.847	0.871	0.352	0.884
ABDE	0.05	1	0.042	1	0.931	0.852	0.107	0.902	0.957
ACDE	0.05	1	0.016	0.993	0.935	0.132	0.869	0.897	0.945
BCDE	0.05	1	0.014	0.966	0.170	0.945	0.930	0.807	0.945
ABCDE	0.2	1	0.259	1	0.907	0.912	0.921	0.879	0.935

Table 6 Theoretical and real weights, theoretical and real characteristic functions, and membership degrees of all players in all possible coalitions for the 2002-2006 Lower House of the Czech Parliament. Codes: ODS=A, CSSD=B, KSCM=C, KDU-CSL=D, SZ=E. Source: own calculations.

The normalized Shapley values of political parties calculated with respect to votes in the 2006-2010 Lower House of the Czech parliament are given in Table 7. The correlation coefficient of the Shapley values with the coefficient of success is 0,92 and is statistically significant at 5% level.

CSSD	ODS	KSCM	KDU-CSL	SZ
0.307	0.327	0.184	0.114	0.068

Table 7 Normalized Shapley value of political parties present in the 2006-2010 Lower House of the Czech Parliament based on the voting outcomes. Source: own calculations.

Expected Shapley value of the 2010-2012 Lower House

In order to predict potential success of players in voting game, we have to take into account their potential power calculated by a-priori power indices, as well as their real voting during precedent periods. In the 2010-2012 Lower House of the Czech Parliament, there were three political parties which were present in two precedent periods (ODS, CSSD and KSCM). Number of players in coalitions was determined using membership degree; the membership degree of two new players (TOP09 and VV) was estimated by the average value membership degree of all players in coalitions. The weights of coalitions were estimated as average values of weights in preceding parliamentary outcome. As players ODS, TOP09 and VV created a-priori coalition, the average weights of coalitions containing these players were adjusted to be the same as the estimated weight ratio in a-priori coalition in 2002-2006 Lower House. Estimated normalized Shapley values as well as Shapley values for a-priori coalition of ODS, TOP09 and VV are given in Table 8. The correlation coefficients are 0,849, and 0,959, respectively. Both correlation coefficients are better comparing to the a-priori Shapley-Shubik indices, even though only the second one is statistically significant at 0.05 level.

	ODS	CSSD	KSCM	KDU-CSL	US
Normalized Shapley value	0.331	0.158	0.055	0.228	0.228
Normalized Shapley value with a-priori coalition of ODS, TOP09, and VV	0.421	0	0	0.289	0.289

Table 8 Expected Normalized Shapley value of political parties present in the 2010-2012 Lower House of the Czech Parliament based on the voting outcomes in precedent periods. Source: own calculations.

4 Conclusion

This article compares the calculated a-priori Shapley-Shubik power index and the Shapley value calculated with respect to obtained votes with the voting success of players in the real voting game – in this case the data from the voting in the Lower House of the Czech Parliament 2002-2012. The calculated a-priori Shapley-Shubik power indices do not correlate with calculated indices of success. For the 2002-2006 Lower house, the correlation coefficient of the index of success with the calculated Shapley-Shubik power index is -0.073; the correlation coefficient of the index of success with the Shapley-Shubik power index with a-priori coalition structure is 0.664. The correlation coefficient of the index of success with the calculated Shapley-Shubik power index for the 2006-2010 Lower House is 0.704. All three correlation coefficients are not statistically significant at 0.05 level. The Shapley values based on real voting outcome improve the correlation – calculated correlation coefficients are higher, they vary from 0.39-0.92.

Obtained data are suitable to determine the membership degree of players – in this case political parties – in all possible coalitions. These values can be used to determine the expected values of characteristic function of the game, and can be used to estimate the real power of players. The results of such estimation evince improvement in the correlation coefficients.

References

[1] Butnariu, D., and Kroupa T.: Shapley mappings and the cumulative value for n-person games with fuzzy coalitions. *European Journal of Operational Research*. Vol. 186, No. 1, 2008, 288-299.

[2] Mareš, M.: *Fuzzy cooperative games*: *cooperation with vague expectations*. New York: Physica-Verlag, 2001.

[3] Mielcova E.: The Uncertainty in Voting Power: The Case of the Czech Parliament 1994-2004. *AUCO Czech Economic Review* Vol. 4, No. 2, 2010, 201–222.

[4] Shapley, L. S.: A value for n-Person Games. In: *Contributions to the Theory of Games*, volume II, (Kuhn H.W. and Tucker A.W., eds.). *Annals of Mathematical Studies* v. 28, 1953, Princeton University Press, 1953, 307–317.

[5] Shapley L. S., and Shubik M.: A Method for Evaluating the Distribution of Power in a Committee System. *The American Political Science Review* Vol. 48, No. 3 (Sep., 1954), 787-792.

[6] Owen, G.: *Game theory*. Academic Press, San Diego, 1995.

[7] Owen, G.: Values of Games with a Priori Unions. In: *Essays in Mathematical Economics and Game Theory* (Hein R. and O. Moeschlin eds.), Berlin: Springer-Verlag, 1977, 76-88.

[8] Turnovec, F.: *Public choice*: *models, methods and applications*. 1st ed. Karolinum, Prague, 2003.

[9] Yu X, Zhang Q.: The Shapley value for fuzzy bi-cooperative games. [online] [accessed: 2012-05-18]. From: www.atlantis-press.com/php/download_paper.php?id=1445.

Resource allocation among academic departments as a coalition game

Hana Mihalčinová[1]

Abstract. Process of resource allocation among academic departments is evaluated by the theory of games. Resource allocation among academic departments is a yearly plan for using available money from budget. In this paper, the theory is applied to the faculty with eight departments for years 2004-2010.
Each academic department chooses ratio between teaching and research indicators. The departments form a coalition to be able to define the rules of resource allocation.
If this theory is applied, it changes the current method of resource allocation.

Keywords: theory of games, coalition games, redistribution system.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The eight academic departments of the faculty solve allocation of resources each year in March. The budget is allocated by given rules. All departments have a certain amount of money available to be used in the beginning of the year till the moment of the distribution of budget happens.

Meanwhile the faculty is expecting final results of the previous year, 50 % of the budget is allocated to the faculty based on the last year results. In fact, the allocation of first 50% of the budget is defined based on results in the year before the last year. The second half of the budget is defined based on the results of the previous year. So, the current year budget is not really based only on the last year results, but also on the results of the year before the last year.

Teaching and research indicators are finally defined during March of each year. All eight academic departments of the faculty vote for the final ratio how the budget is divided between teaching and research indicators. Each academic department has one voice. Absolute majority has to agree with this setting. The ratio of the indicators can change during the next year though.

In this paper, we assume that the voting happens in March of each year for the rest of current year and for the first three months of next year.

At the beginning of the first year departments voted for 30 % for research indicators and for 70 % for teaching indicators. It is possible that the ratio changes in the coming year, but the change cannot be higher than 10 % for one of the indicators. So the division of the budget in coming year can end up as 40 % for research indicators and 60 % for teaching indicators. Or it can move the other way and the budget can be divided as 20 % for research indicators and 80 % for reaching indicators. This rule of maximum change by 10% in any direction is needed as it would be possible to open the door to the option of accepting as big change as even having 0% of the budget for one of the indicators. Academic departments can form coalition and change indicators easily that way as the distribution of ratio between indicators depends on coalition formed by academic departments. This problem is solved by theory of game and the theory of redistribution systems.

2 Theory

2.1 Theory of Games

The participant of this conflict, the departments, we call *players*. We have eight players A, B, C, D, E, F, G, and H. The players are intelligent. They know details about performance of each other from the last year. Each player votes for ratio between research and teaching indicators according to his last year performance. This is his *strat-*

[1] University of Economics Prague, Department of Econometrics, W. Churchill Sq. 4, 13067 Prague 3, Czech Republic, vysh01@vse.cz

egy. The list of all decisions is called the *strategy spaces*. The allocated money to each department is called *payoff*. And the whole process of decision is called a *game*.

According to properties of strategy spaces and payoffs, we specify this game as *finite game* (all strategy spaces have a finite number of elements) with *constant sum games*.

For simplification, the departments have budget 10 million CZK. The table 1 shows all strategy spaces for all departments. The initial ratio between research and teaching indicators is 30 and 70 % in 2004.

Research indicator	Teaching indicator	A	B	C	D	E	F	G	H	Budget
0%	100%	376	1 052	249	2 292	928	2 125	1 169	1 809	10 000
10%	90%	375	1 110	281	2 278	895	2 057	1 171	1 832	10 000
20%	80%	374	1 168	314	2 264	861	1 989	1 174	1 855	10 000
30%	**70%**	**374**	**1 226**	**347**	**2 250**	**827**	**1 921**	**1 176**	**1 879**	**10 000**
40%	60%	373	1 284	380	2 237	794	1 853	1 178	1 902	10 000
50%	50%	372	1 342	413	2 223	760	1 785	1 180	1 925	10 000
60%	40%	371	1 401	445	2 209	726	1 716	1 183	1 949	10 000
70%	30%	370	1 459	478	2 195	693	1 648	1 185	1 972	10 000
80%	20%	369	1 517	511	2 181	659	1 580	1 187	1 996	10 000
90%	10%	369	1 575	544	2 167	625	1 512	1 189	2 019	10 000
100%	0%	368	1 633	577	2 153	592	1 444	1 192	2 042	10 000

Table 1 Strategy spaces in 2004 (in thousand CZK)

To be able to apply any of the strategies, each player needs to become a member of a coalition.

The *coalition* is a group of players who negotiate which strategy shall be chosen as the goal is clear – to improve their payoffs. One coalition is formed with the aim to focus on decreasing research indicator, and the second coalition focuses on increasing research indicator. In our case, we have eight players, so the winning coalition must have at least five players. The change or research indicator can reach only 10 % in both directions.

Every player wants to maximize his payoff. At the same time each player knows that he can receive more money as long as the other player loses the same amount in favor of the other player.

In the shadow cells, there are shown two possible coalitions. The first coalition is composed of the following departments A, D, E and F. This coalition wants to enforce a reduction of research indicator by 10 %. The second coalition is composed of the remaining departments B, C, G, and H. This coalition wants to enforce a reduction of teaching indicator by 10 %.

In the figure 1, you can see the strategy spaces and payoffs. The initial ratio between research and teaching indicators is shown as vertical axis. From this figure, it is hard to say which strategy to choose. It is hard to define how many players are for increase of research indicator and how many of the players are for its decrease.

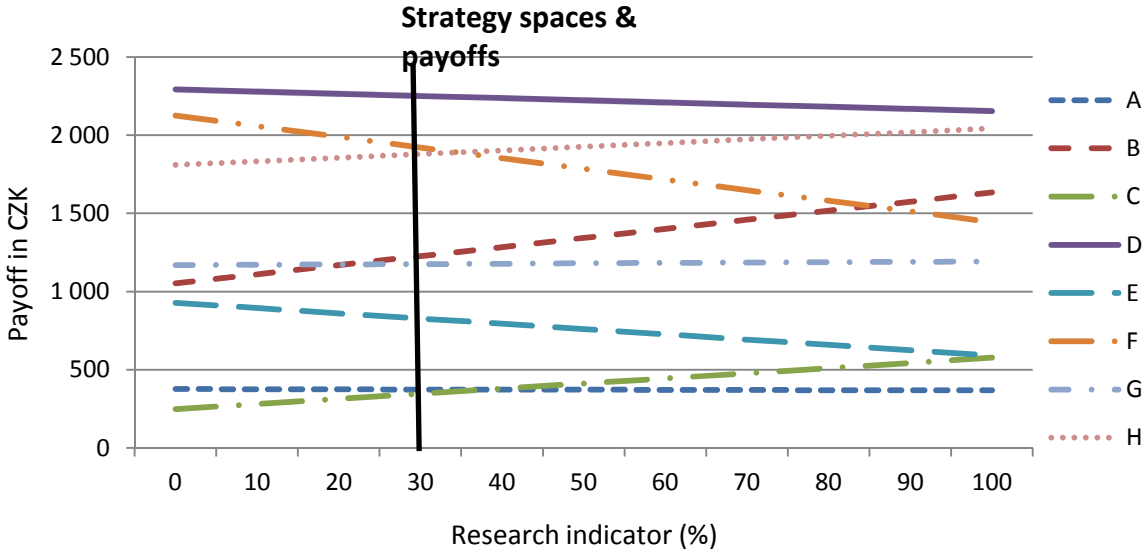


Figure 1 Strategy spaces & payoffs

Therefore it is better to count only with difference between maximum and minimum payoff of departments. In the figure 2 we can see strategy spaces and payoffs for all departments. For example, player F has a payoff higher by 700 thousand, so it is more advantageous to vote for 0 % for research indicator than if he votes for the option of having 100 % for research indicator. This payoff is gotten by F player at the expenses of the players B, C, G, and H though. In this figure 2 it is very good noticeable which departments form a coalition. On the left side from the horizontal axis (30 % research indicator) departments A, D, E and F form the first coalition. This coalition wants to enforce decreasing of research indicator by 10 %. The remaining departments B, C, G, and H form the second coalition and want to enforce an increase of research indicator by 10 %.

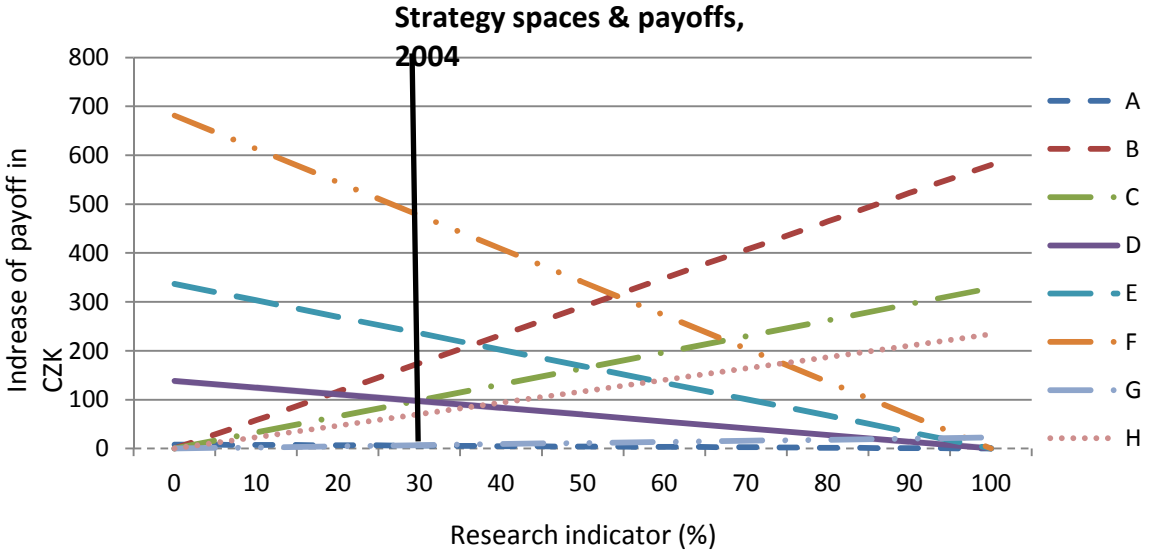


Figure 1 Strategy spaces for increases of money

2.2 Redistribution system

Elementary redistribution system

Elementary redistribution model (ERS) is a special model of game theory, which is a game with more than two players, with free disjunctive coalition structure and with non-constant payoffs.

ERS introduces a game, in which the rules for the amount of payoffs are based on forthcoming assumptions:

- system has only three players (A,B,C)
- outputs of players are 6,4,2
- every player has the same influence strength equal to 1
- all players know their own performance and performance of other players
- all coalitions are possible and equal

It is only needed to fill into the redistribution equation according to which the payoffs are redistributed:

$$x + y + z = 12 - \eta * R(x - 6, y - 4, z - 2)\,, \tag{1}$$

where $x+y+z$ is a sum of real payoff of player A,B, and C

12 is a maximum of payoff which can be redistributed if the maximum performance is reached

η is a coeficient of how much the performance will be lowered

R(x-6,y-4,z-2) is a function of distance of redistribution of real payoffs from payoffs according to performance which we assume:

$$R(x - 6, y - 4, z - 2) = \sqrt{(x-6)^2 + (y-4)^2 + (z-2)^2}\,, \tag{2}$$

In this system, we do not work with the performance reduction as it is very difficult to estimate the lost. It is impossible for us to calculate the decrease in performance of academic department.

2.3 Redistribution system applied to eight academic departments

In the text above, we read about the strategy spaces and the payoffs for all departments in 2004. We also notice possible coalitions for the next year 2005. The process of forming of coalition will be described now.

In 2004, the player with the highest performance, player F, wants to maximize his payoff for the next year. The advantage for him is that he knows how much money can be gotten by decreasing the research indicator by 10 % at the expenses of other players and also which players will lose the payoff at the same time.

The player F forms a coalition with other players who can get more money by decreasing research indicator by 10%. The second coalition votes exactly for the opposite, which is the increase of the same indicator by 10 %.

The coalition is formed by 4 players. As each coalition has the same number of members, the final result of voting stays the same as it was before the voting process started or we can applied redistribution system.

The player *A* is coalition formed by player F. The player *B* is the second coalition opposite the coalition formed by F. In this coalition is also the player G with the smallest increment. The player *C* is the player G. The player F can get 68 thousand from budget and player C can lose only 4 thousand by voting for 20 % research indicator. The player F offers to player G at least this amount, 4 thousand and player G will vote for 20% research indicator. Player F divides his increment.

In the following year, a new coalition with 5 members will be formed to act against the second coalition with only 3 members. So, the result will not be the same. The research indicator will be increased by 10 percent then. The situation for 2005 is shown in the table 2.

Research indicator	Teaching indicator	A	B	C	D	E	F	G	H	Budget
0%	100%	346	1 039	211	2 402	900	2 170	1 132	1 801	10 000
10%	90%	347	1 102	244	2 356	867	2 105	1 146	1 833	10 000
20%	80%	347	1 165	278	2 310	835	2 040	1 160	1 865	10 000
30%	**70%**	**348**	**1 229**	**311**	**2 264**	**803**	**1 975**	**1 174**	**1 897**	**10 000**
40%	60%	348	1 292	345	2 218	770	1 910	1 188	1 930	10 000
50%	50%	349	1 355	378	2 171	738	1 845	1 202	1 962	10 000
60%	40%	349	1 418	411	2 125	706	1 781	1 216	1 994	10 000
70%	30%	350	1 482	445	2 079	673	1 716	1 230	2 026	10 000
80%	20%	350	1 545	478	2 033	641	1 651	1 243	2 059	10 000
90%	10%	350	1 608	511	1 987	609	1 586	1 257	2 091	10 000
100%	0%	351	1 672	545	1 941	576	1 521	1 271	2 123	10 000

Table 2 Strategy spaces in 2005 (in thousand CZK)

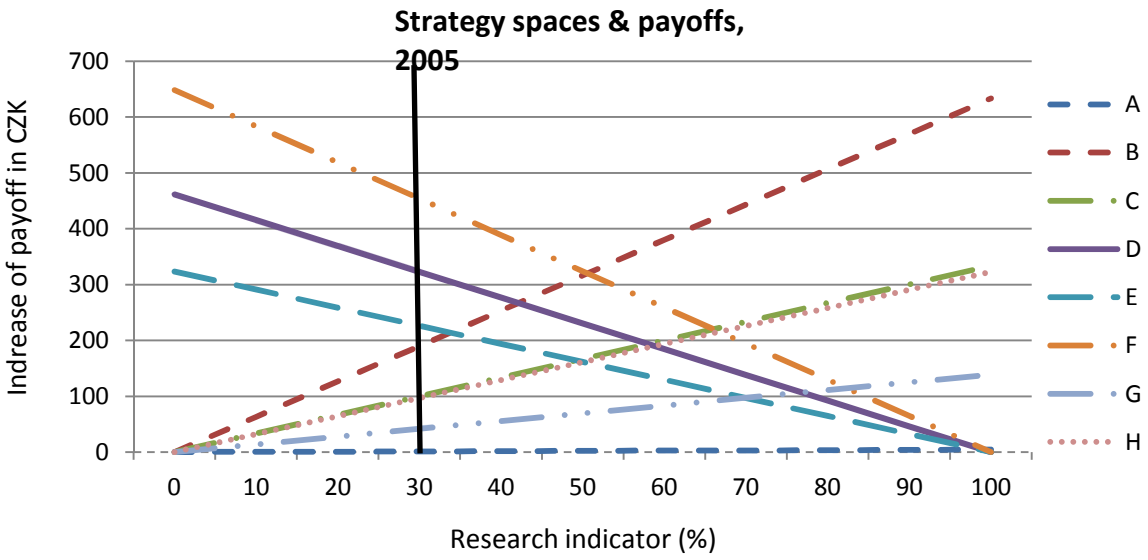


Figure 1 Strategy spaces for increases of money

3 Results for 2004 - 2010

The table shows the division of the budget among departments of the faculty. The whole budget is % divided to academic department in 2010.

Distribution of budget	A	B	C	D	E	F	G	H	Budget
	4%	17%	2%	24%	9%	11%	13%	19%	100%

Table 3 Distribution of budget in 2010 (in %)

The table below shows the changes in division of the budget among the departments of the faculty in 6 years. How much better/worse off the individual academic departments for the year 2004-2010.

Changes in 6 years	A	B	C	D	E	F	G	H
Absolute change	35	5	-5	-61	20	14	12	-19

Table 4 Total changes of budget for 2004 - 2010 (in thousand CZK)

4 Conclusion

If player G in 2004 will proceed to the offer from player F, the whole faculty can record decline of the whole faculty. So, if somebody is the weakest player, with the minimal increment, is the player *C* from redistribution system. This player causes a decrease in the whole system.

Acknowledgements

This research was funded by the Internal Grant Agency or the University of Economics Prague, project no. F4/16/2011 Models of Operational and Financial Management.

References

[1] Benesch, J.: Diplomová práce. *Teorie her a ekonomické systémy*. Vysoká škola finanční a správní o.p.s., Praha, 2008.
[2] Benesch, J.: Co brání meziorganizační migraci?, Marathon, 2007.
[3] Blackwell, D.; Girshick, M,A,: *Teorie her a statistického rozhodování*, Knihtisk Praha, 1964.
[4] Carmichael, F.: *A Guide to Game Theory*, Person Education Limited, Harlow, 2003.
[5] Fiala, P.: *Modely a metody rozhodování*. Vysoká škola ekonomická v Praze, 2003.
[6] Fisher, R., Ury, W., Patton, B.: *Jak dosáhnout souhlasu*. Management Press, Praha, 2006.
[7] Maňas, M.: *Teorie her a konflikty zájmů*. Vysoká škola ekonomická v Praze, 2002.
[8] Mihalčinová, H.: Conditions which influence performance of economic system from perspective of game theory. *Mathematical methods in economics* (2010), 441–445.
[9] Montet, Ch.; Serra, D.: *Game theory and economics*, Palgrave Maxmilián, Houndmills- GB, 2003.
[10] Osborne, J.: *An Introduction to Game Theory.* University Press, Oxford, New York, 2004.
[11] Plamínek, J.: *Konflikty a vyjednávání*, Grada Publishing, a.s., Praha, 2009
[12] Selten, R.: *Game theory and behavior: selected essays, Volume 1*, Cheltenham-GB, Elgar, 1999.
[13] Selten, R.: *Game theory and behavior: selected essays, Volume 2*, Cheltenham-GB, Elgar, 1999.
[14] Valenčík, R.: *Základy terie redistribučních systémů a její aplikace*, Vysoká škola finanční a správní o.p.s., Praha, 2008.

Risk of abrupt changes in the property rights protection

Štěpán Mikula [1]

Abstract. The numerous researches identify the level of property rights protection as a key determinant of economic performance. My previous research has also shown that the presence of abrupt changes in development of property rights protection also matters. This paper aims to explain probability of occurrence of these abrupt changes by properties of political systems.

Presence of abrupt changes is evaluated using techniques of measuring predictability of institutional environment based on modelling evolution of institutions as ARI(p,d) processes. The alternative way relying on detection of outliers in time-series of growth rates is also presented. The estimated probability of occurrence of abrupt changes is then explained by quality and stability of political systems.

Keywords: institutions, predictability, ARI process, abrupt changes, property rights

JEL classification: P48
AMS classification: 91B84

1 Introduction

There is a broad consensus in recent literature on importance of institutional environment for economy and its performance. Empirical research has shown that the protection of property rights is the most important part of the general bundle of institutional environment [1]. These researches are mostly focused on the state of property rights protection. However as it could be seen in the figure 1a the average level of property rights protection can be almost identical (Cameroon: $\bar{x} = 4.76$; Burkina Faso: $\bar{x} = 4.83$) but the way of development in time may differ significantly.

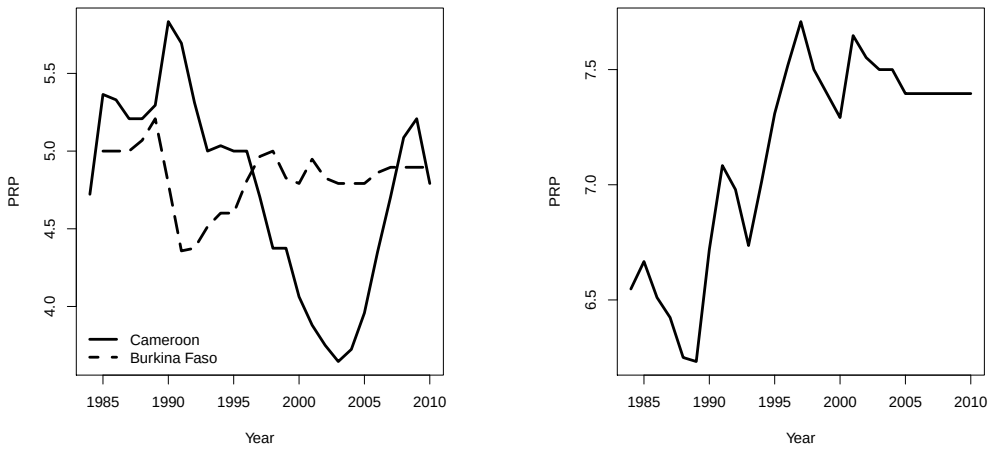
My previously presented results [6] indicate, that the way of development matters with respect to economic performance as well as the level of protection. The way of development was investigated using preliminary version of techniques of evaluation of institutional environment predictability (more advanced version is described in section 3.1). The greatest importance with respect to performance was found in the case of negative abrupt (i.e. unexpected) changes. Here follows the natural question: What determines whether the abrupt changes occurs?

This paper deals with this question. This paper assumes that the main determinant of occurrence of abrupt changes should be in politics. It is a logical consequence of the fact that property rights setting is clearly issue of politics. If the government is able to set certain level of property rights protection then is also able to change it abruptly. Therefore this paper investigates whether political system matters with respect to occurrence of abrupt changes.

2 Data

As I have already mentioned the property rights protection is one of the most important parts of broader set of institutions or institutional environment. Institutions do not favour empirical research because they are directly unmeasurable. Therefore it is necessary to rely on proxies which are often based on expert evaluation. The level of property rights protection is commonly approximated by index originating from

[1] Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, Brno, stepan.mikula@gmail.com



(a) PRP in Cameroon and Burkina Faso (b) PRP in Czech Republic

Figure 1: Development of PRP in selected countries during 1984–2010

Knack & Keefer [3] which is based on expert evaluations from International Country Risk Guide (ICRG). Its value is obtained as follows:

$$PRP = \frac{10}{16}BQ + \frac{10}{24}Cor + \frac{10}{48}IP + \frac{10}{24}LaO \tag{1}$$

Where BQ is Bureaucracy Quality, Cor Corruption, IP Investment Profile and LaO Law and Order (for further definition see [7]). Used weights equalize weights set in ICRG and scale resulting index on the continuous scale from 0 to 10 where higher value of PRP means better level of property rights protection. PRP index is available on yearly basis for 140 countries since 1984. This limits analysed period to 1984–2010.

Data for transitional economies of former eastern block are shortened to period 1994–2010 because it is reasonable to expect structural changes in the beginning of nineties.

3 Detection of abrupt changes

At first it is necessary to analyse time-series of PRP index to determine whether abrupt changes are present. This paper uses two approaches to this task. The first one tries to evaluate overall predictability of development of time-series which is naturally decreased by presence of abrupt changes. The second approach is focused on searching for specific observations which can be classified as abrupt changes in property rights protection. Both approaches are described in detail in the following sections.

3.1 Predictability

Concept of measuring predictability of institutional environment is based on the theory of path-dependent way of evolution of institutions. This widely accepted theory assumes that current state of institutions is not random but it depends on its past states. This assumption allows to describe development of institutional environment as ARI(p,d) process. It means that it is also reasonable to presume that agents makes their expectations on the basis of past states. Predictions which come from estimated ARI(p,d) process therefore describes expectations of agents. Basically this approach works with agents with adaptive learning.

There are almost certainly more predictors of development of property rights which agents take into account (e.g. risk of conflicts, revolutions, coup d'états, state bankruptcy etc.), but these predictors may significantly differ around the world. Moreover these additional predictors probably differ for various aspects of institutional environment. On the other hand it is necessary to keep in mind that method based solely on ARI(p,d) processes almost certainly undervalue the real predictability of institutional environment.

Evaluation of predictability is carried out in the following steps:

1. Time-series of PRP index is tested for stationarity by Augmented Dickey-Fuller test ($\alpha = 10\,\%$). If the time-series is not stationary (time-series usually contains stochastic trend – see figure 1b) then the time-series is differentiated till it is stationary. Basically in this step the parameter d of an ARI(p,d) process is found.

2. Parameters of ARI(p,d) process are estimated for $p \in [0, 10]$ using maximum-likelihood estimator and one-step ahead in-sample predictions are made using Kalman filter.

   This step needs further explanation. Agents certainly take into account more future values than only one. However results that come from one-step ahead predictions and n-steps predictions where $n \in [2, 5]$ are well correlated.[1] Using in-sample predictions is also questionable, because it basically assumes that agents know their own future – which is of course a little unrealistic. On the other hand using in-sample predictions allows evaluation of time-series without their serious shortening, which is necessary if recursive techniques are used. Moreover results that comes from in-sample and out-of-sample predictions are again well correlated.

3. Goodness-of-fit is evaluated for every p using standard defined RMSE statistics:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n}} \tag{2}$$

4. Model which minimize RMSE is chosen.

5. On the basis of chosen model are calculated statistics RRMSE and RRMSNE defined by equations (3), (4) and (5). These statistics are then used as measures of predictability. Note that smaller values of RRMSE and RRMSNE means better predictability.

$$RRMSE = \sqrt{\frac{\sum_{t=1}^{n}\left(\frac{y_t - \hat{y}_t}{y_t}\right)^2}{n}} \tag{3}$$

$$RRMSNE = \sqrt{\frac{\sum_{t=1}^{n} e_t^2}{n}} \tag{4}$$

$$e_t = \begin{cases} \frac{y_t - \hat{y}_t}{y_t} & \text{if } (y_t - \hat{y}_t) < 0 \\ 0 & \text{if } (y_t - \hat{y}_t) \geq 0 \end{cases} \tag{5}$$

   These statistics originates from common RMSE. However they alter RMSE to reflect some specific aspects. RRMSE assigns different weights to the same absolute difference between fitted and observed variable which occurs under different state of PRP. If the level of property rights protection is low then the same absolute difference has bigger importance than if it would be high and vice versa. It makes sense – the same absolute difference may result in total expropriation under weak property rights protection and in just a little hiccup when PRP is high.

   RRMSNE takes into account only these observation where fitted value is bigger than observed value – i.e. situations when agents have expected better protection than has in fact occurred. It is possible to understand RRMSNE as a measure of unpleasant surprises. RRMSNE is important because of previously mentioned results that emphasize importance of unexpected negative changes in PRP.

   It is reasonable to expect that there should be close relationship between RRMSE and RRMSNE. This is indeed true – as it could be seen in the figure 2. However the closer look on their dependence reveals that for higher values of both statistics (i.e. worse predictability) their relation become significantly weaker. This is caused by presence of countries which experienced huge negative abrupt changes in the level of the property rights protection.

3.2 Growth rates

The second approach to detection of abrupt changes is much more simple. It focuses on identification of abrupt negative changes in time-series of PRP and therefore is closer to RRMSNE than to RRMSE. This

[1] Predictions for $n > 5$ were not tested, however there is no reason to expect any change.

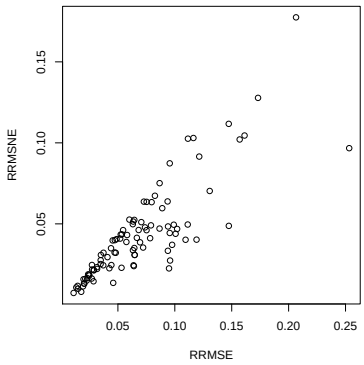


Figure 2: Correlation of RRMSE and RRMSNE with removed outliers

approach is also more oriented to concrete events compared to statistics of predictability which evaluates overall predictability of whole time-series.

As unexpected is considered such change which is bigger then usual. Evaluation is performed for each time-series. At first the annual growth rates are calculated and then the possible outliers are found. Observation y_t which meets condition (7) is considered to be an outlier. The number of identified outliers GRO can be taken as approximation of probability of abrupt negative change occurrence.

$$GRO = \sum_{t=1}^{n} x_t \tag{6}$$

$$x_t = \begin{cases} 1 & \text{if } y_t < \bar{y} - 2\sqrt{\frac{(y_t-\bar{y})^2}{n-1}} \\ 0 & \text{if } y_t \geq \bar{y} - 2\sqrt{\frac{(y_t-\bar{y})^2}{n-1}} \end{cases} \tag{7}$$

Results of analysis are depicted in thematic map in figure 3a. These results also reveal drawbacks of this approach. According this evaluation some countries (e.g. USA, Germany or Canada) experienced non-zero number of negative abrupt changes – even though their biggest negative growth rates are small in comparison to another more volatile countries.

This drawback can be partially corrected by using standard deviation obtained from all growth rates of all countries. This modification is used for calculation of GRO_w. Distribution of GRO_w is depicted in the figure 3b.

(a) GRO (b) GRO_w

Figure 3: Distribution of GRO and GRO_w around the world.

4 Explanation of abrupt changes

4.1 Explanatory variables

This paper seeks explanation of occurrence of abrupt changes in features of political system. It focuses on two characteristics: quality and stability. Quality of political system is described using following variables:

polity2 is variable from dataset Polity IV [4] which evaluate nature of regime on the discrete scale from -10 to 10, where -10 is extreme value for autocracy and 10 extreme for democracy. Values are available on yearly basis. I used average values for whole period which can be treated as continuous in value.

left originates from dataset DPI2010 [2] (variable *execrlc*) and reflects the probability that a left-wing party is in power. Variable $left$ is computed as follows:

$$left = \frac{1}{n}\sum_{t=1}^{n} x_t \tag{8}$$

$$x_t = \begin{cases} 1 & \text{if left-wing party is in power} \\ 0 & \text{if left-wing party is not in power} \end{cases} \tag{9}$$

Variable $left$ is included because of assumed tendency of left-wing parties to restrict private property.

allhouse also comes from DPI2010 and indicates whether the party in power controls all relevant houses. Value of *allhouse* is computed analogously to the $left$.

maj is last used variable from DPI2010 and reflects average fraction of seats held by the government. This variable, as well as *allhouse*, reflects ability of government to carry out the policy that it choose to implement.

Additional variable describes stability of political system:

durable is average number of years between change of *polity2* of at least 3 points. Variable *durable* comes from Polity IV.

4.2 Model

Design of the model has to deal with collinearity of variables which describe quality of political systems. This issue is solved by principal component analysis (PCA) performed on scaled matrix of these variables. First two extracted orthogonal components are used as explanatory variables. These two components account for 76.6 % of observed variability. Correlation of used components and original variables is depicted in table 1.

	polity2	*left*	*allhouse*	*maj*
Q_{pc1}	-0.85	0.15	0.69	0.88
Q_{pc2}	0.32	0.93	0.31	-0.10

Table 1: Correlation (Pearson's ρ) between PCA scores and variables describing quality of political system

The econometric model for explaining RRMSE and RRMSNE has therefore following form:

$$\log(RRMSE) \text{ or } \log(RRMSNE) = \beta_0 + \beta_1 Q_{pc1} + \beta_2 Q_{pc2} + \beta_3 \log(durable) \tag{10}$$

This model is estimated on cross-sectional data ($n = 125$) using OLS and the results are presented in table 2.

The model for explaining GRO and GRO_w has altered structure. There is an additional variable which describes stability of political system – *coup*. This variable [5] is the number of successful and attempted coups d'état during examined period 1984–2010. Number of coups is added because of event-oriented nature of GRO. Altered model has the following structure:

$$GRO \text{ or } GRO_w = \beta_0 + \beta_1 Q_{pc1} + \beta_2 Q_{pc2} + \beta_3 \log(durable) + \beta_4 coups \tag{11}$$

Equation (11) was estimated on the same data and by the same estimator as (10). Results are also presented in table 2. The model for GRO was quite unsurprisingly found insignificant and therefore is not included in the table.

Explained var.	*const.*	Q_{pc1}	Q_{pc2}	$\log(durable)$	*coups*	$\bar{R}^2$
$\log(RRMSE)$	-1.400^{***} (0.189)	0.112^{***} (0.043)	-0.116^{*} (0.060)	-0.422^{***} (0.064)		0.34
$\log(RRMSNE)$	-2.237^{***} (0.178)	0.114^{***} (0.040)	-0.119^{**} (0.056)	-0.324^{***} (0.059)		0.30
GRO_w	1.034^{***} (0.230)	0.007 (0.050)	-0.139^{**} (0.068)	-0.243^{***} (0.074)	0.1312^{***} (0.044)	0.22

Table 2: Results of OLS regression

4.3 Results

Estimated results shows that there is indeed relationship between occurrence of abrupt changes and political system. This relationship is clear especially for variables which reflects stability of political system. They are in all cases significant with expected values. Better stability (higher *durable* and lower *coup*) decrease probability of occurrence of abrupt changes – i.e. decrease of all dependent variables.

Relationship between quality and dependent variables seems to be much more complicated. The first component Q_{pc1} meets the expectations. However it is significant just in the case of RRMSE/RRMSNE. The value of estimated parameters for the second one which is closely correlated to the $left$ is inverse to expectations. It might be caused by suboptimal selection of variables or by actual insignificance of the $left$. This issue however requires further research.

5 Conclusion

This paper has investigated whether features of political systems determines probability of occurrence of abrupt changes in property rights protection level. Using cross-sectional data the significant dependence was found on the stability and quality of political systems. These preliminary results seems to promising for next research which should be focused on two areas: a) improvement in methodology of measuring of predictability (including analysis of wider set of indicators) and b) augmentation of models (10) and (11) with broader set of control variables.

Acknowledgements

I am grateful to MetaCentrum for providing sources for development of methods of measuring predictability of institutional environment. This research is supported by grant no. MUNI/A/0786/2011.

References

[1] Acemoglu, D., and Johnson, S.: Unbundling Institutions. *Journal of Political Economy* **113** (2005), 949–995.

[2] Keefer, P.: *Database of Political Institutions DPI2010,* on-line database, World Bank, 2010.

[3] Knack, S., and Keefer, P.: Institutions and Economic Performance: Cross-Country Test Using Alternative Institutional Measures. *Economics and Politics* **7** (1995).

[4] Marshall, M. G., Jaggers, K., and Gurr, T. R.: *Polity IV Project: Political Regime Characteristics and Transitions, 1800–2010* (2011), on-line database.

[5] Marshall, M. G., Marshall, D., R.: *Coup d'État Events, 1946–2010* (2010), on-line database.

[6] Mikula, Š.: On the Predictability of Institutional Environment. In: *Mathematical Methods in Economics 2011* (M. Dlouhý and V. Skočdopolová, eds.), University of Economics, Prague, 2011.

[7] PRS Group: *International Country Risk Guide (ICRG)* (2012), on-line available at: `https://www.prsgroup.com/ICRG.aspx`.

Efficient algorithm for checking periodicity of interval circulant fuzzy matrices

M. Molnárová[1], H. Myšková[2], J. Plavka[3]

Abstract. Periodic properties of circulant interval matrices over $(\max, \min)$-algebra (fuzzy matrices) are studied. Necessary and sufficient conditions for possible and universal d-periodicity of circulant interval matrices are proved. $O(n)$ algorithm for verifying the possible d-periodicity and another $O(n \log n)$ algorithm for verifying the universal d-periodicity as well are described.

Keywords: (max, min) algebra, interval matrix period, circulant matrix

JEL classification: C02
AMS classification: 08A72, 90B35, 90C47

1 Introduction

To study matrix properties in $(\max, \min)$-algebra, where addition and multiplication are formally replaced by operations of maximum and minimum, is of great importance for applications in various areas. Periodic behaviour of fuzzy matrices with corresponding polynomial algorithms were studied in [3] and [6]. However, in practice we deal often with inexact input data. This leads to demand replace scalar matrices by so-called interval matrices ([1]).

2 Preliminaries

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 and by $\mathbb{N}_0$ the set $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. The greatest common divisor of a set $S \subseteq \mathbb{N}$ is denoted by $\gcd S$. 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 $n \in \mathbb{N}$, $B(n, n)$ denotes the set of all square matrices of order n and $B(n)$ the set of all n-dimensional column vectors over $\mathcal{B}$. Matrix operations over $\mathcal{B}$ are defined formally in the same manner (with respect to $\oplus$, $\otimes$) as matrix operations over any other field. The rth power of a matrix A is denoted by A^r, with elements $a_{ij}^{(r)}$. For $A, C \in B(n, n)$ we write $A \leq C$ ($A < C$) if the inequality $a_{ij} \leq c_{ij}$ ($a_{ij} < c_{ij}$) holds for all $i, j \in N$.

For a matrix $A \in B(n, n)$ the symbol $G(A) = (N, E_G)$ stands for a complete, arc-weighted digraph associated with A, i.e., the node set of $G(A)$ is N, and the capacity of any arc (i, j) is a_{ij}. Let $\emptyset \neq \tilde{N} \subset N$. $G/\tilde{N}$ stands for a subdigraph of digraph $G(A) = (N, E_G)$ with the node set $\tilde{N}$ and arc set $E_{G/\tilde{N}} = \{(i, j) \in E_G;\ i, j \in \tilde{N}\}$. A path in the digraph $G(A) = (N, E_G)$ is a sequence of nodes $p = (i_1, \ldots, i_{k+1})$ such that $(i_j, i_{j+1}) \in E_G$ 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.

By a *strongly connected component* $\mathcal{K}$ of $G(A) = (N, E_G)$ we mean a subdigraph $\mathcal{K}$ generated by a non-empty subset $K \subseteq N$ such that any two distinct nodes $i, j \in K$ are contained in a common cycle

[1]Technical University of Košice, Department of Mathematics and Theoretical Informatics, B. Němcovej 32, 04200 Košice, Slovakia, e-mail: Monika.Molnarova@tuke.sk
[2]Technical University of Košice, Department of Mathematics and Theoretical Informatics, B. Němcovej 32, 04200 Košice, Slovakia, e-mail: Helena.Myskova@tuke.sk
[3]Technical University of Košice, Department of Mathematics and Theoretical Informatics, B. Němcovej 32, 04200 Košice, Slovakia, e-mail: Jan.Plavka@tuke.sk

and K is a 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. The set of all strongly connected components of G is denoted by $\mathrm{SCC}(G)$.

Definition 1. Let $A \in B(n,n)$. The *matrix period*, in notation $\operatorname{per} A$, is defined as the minimal natural number p for which there is $R \in \mathbb{N}$ such that

$$A^{k+p} = A^k \quad \text{for all} \quad k \geq R.$$

In a $(\max,\min)$-algebra any element of any power of the matrix A is equal to some element of A. Therefore, the power sequence of A contains finitely many different matrices with entries of A only. As a consequence, a fuzzy matrix is always periodic in contrast to matrices in another extremal algebra, namely, max-algebra.

For given $h \in B$, the *threshold digraph* $G(A,h)$ is the digraph with the node set N and with the arc set $E_G = \{(i,j);\ i,j \in N,\ a_{ij} \geq h\}$.

The following lemma describes the relation between matrices and corresponding threshold digraphs.

Lemma 1. [4] 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 $G(A,h_2) \subseteq G(A,h_1)$.

Following theorems proved in [3] are useful for study periodic properties of interval matrices.

Theorem 1. [3] *Let $A \in B(n,n)$, $d \in \mathbb{N}$. Then the following assertions are equivalent*

i) $\operatorname{per} A | d$,

ii) $(\forall h \in B)(\forall \mathcal{K} \in \mathrm{SCC}^\star(G(A,h)))\operatorname{per}\mathcal{K} | d$.

Theorem 2. [3] *Let $A \in B(n,n)$. Then*

$$\operatorname{per} A = \operatorname{lcm}\{\operatorname{per}\mathcal{K}; \mathcal{K} \in \mathrm{SCC}^\star(A)\}$$

3 Periodicity of interval circulant matrices

In this section we present a necessary and sufficient condition for an interval circulant matrix to be possibly d-periodic and a necessary and sufficient condition for an interval circulant matrix to be universally d-periodic as well. In addition we describe an $O(n)$ algorithm for verifying the possible d-periodicity and another $O(n \log n)$ algorithm for verifying the universal d-periodicity of an interval circulant matrix.

Definition 2. A matrix $A \in B(n,n)$ is called *circulant*, if it has the form

$$A = \begin{pmatrix} a_1 & a_2 & a_3 & \dots & a_{n-1} & a_n \\ a_n & a_1 & a_2 & \dots & a_{n-2} & a_{n-1} \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ a_2 & a_3 & a_4 & \dots & a_n & a_1 \end{pmatrix}.$$

We denote a circulant matrix A by abbreviation $A(a_1,\dots,a_n)$. The set of all entries of A with the same index is called *a stripe*, the entries a_k form the kth stripe. In the associated digraph $G(A)$ each stripe k defines a set of arcs of the form $(i, i+k-1)$ for $i = 1,2,\dots,n$; obviously all the numbers here are considered modulo n. Denote by $E^{(k)}$ the set of all arcs (i,j) in $G(A)$ corresponding to the kth strip.

We shall say that the *span* of an arc in $E^{(k)}$ is $s_k = k - 1$. As was observed already in [2], the arcs of $E^{(k)}$ fall into a set of disjoint cycles, all with the same length equal to 1 for the first stripe and equal to

$$\ell(c_k) = \frac{n}{\gcd(n, s_k)} \tag{1}$$

for $k \neq 1$. Denote by $C^{(k)}$ the set of all these cycles.

As a consequence results proved in lemmas below can be obtained, which allows to derive the formula for computing the period of a circulant matrix in Theorem 3.

Lemma 2. [4] Let $A(a_1, \ldots, a_n)$ be a circulant matrix. For each $h \in \{a_i; i \in N\}$ the threshold digraph $G(A, h)$ is either strongly connected or consists of m strongly connected components isomorphic with $\mathcal{K}^1$, where $\mathcal{K}^1 \in \mathrm{SCC}^\star(G(A, h))$ containing node 1 and $m \mid n$.

Lemma 3. Let $G' \subseteq G$, $\mathcal{K} \in \mathrm{SCC}^\star(G)$ and $\mathcal{K}' \in \mathrm{SCC}^\star(G'/N_{\mathcal{K}})$. Then $\operatorname{per}\mathcal{K} | \operatorname{per}\mathcal{K}'$.

Proof. Since $\{\, \ell(c);\ c \text{ is a cycle from } \mathcal{K}' \,\} \subseteq \{\, \ell(\tilde{c});\ \tilde{c} \text{ is a cycle from } \mathcal{K} \,\}$ we obtain

$$\operatorname{per}\mathcal{K} = \gcd\{\, \ell(c);\ c \text{ is a cycle from } \mathcal{K} \,\} | \gcd\{\, \ell(\tilde{c});\ \tilde{c} \text{ is a cycle from } \mathcal{K}' \,\} = \operatorname{per}\mathcal{K}'.$$

□

Theorem 3. *Let $A(a_1, \ldots, a_n)$ be a circulant matrix. Then for the period of matrix A holds*

$$\operatorname{per} A = \operatorname{per}\mathcal{K}^1$$

where $\mathcal{K}^1 \in \mathrm{SCC}^\star(G(A, \max\limits_{i \in N} a_i))$ containing node 1.

Proof. According to Lemma 2, Lemma 3 and $G(A, \max\limits_{i \in N} a_i) \subseteq G(A, h)$ for each $h \in \{a_k;\ k \in N\}$ we get $\operatorname{per}\mathcal{K} \mid \operatorname{per}\mathcal{K}^1$ for each $\mathcal{K} \in \mathrm{SCC}^\star(A)$. Since $\mathcal{K}^1 \in \mathrm{SCC}^\star(A)$, we have

$$\operatorname{per} A \ = \ \operatorname{lcm}\{\, \operatorname{per}\mathcal{K}; \mathcal{K} \in \mathrm{SCC}^\star(A)\} = \operatorname{per}\mathcal{K}^1.$$

□

In this paper we shall deal with matrices with interval elements. Similarly to [1], [4], [5] we define an interval matrix $\mathbf{A}$ as follows.

Definition 3. Let $\underline{A}, \overline{A} \in B(n, n)$, $\underline{A} \leq \overline{A}$. *An interval matrix* $\mathbf{A}$ with bounds $\underline{A}$ and $\overline{A}$ is defined as follows

$$\mathbf{A} = [\underline{A}, \overline{A}] = \left\{\, A \in B(n, n);\ \underline{A} \leq A \leq \overline{A} \,\right\}.$$

Definition 4. For an interval matrix $\mathbf{A}$ of the form

$$\mathbf{A} = \begin{pmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3 & \ldots & \boldsymbol{a}_{n-1} & \boldsymbol{a}_n \\ \boldsymbol{a}_n & \boldsymbol{a}_1 & \boldsymbol{a}_2 & \ldots & \boldsymbol{a}_{n-2} & \boldsymbol{a}_{n-1} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \boldsymbol{a}_2 & \boldsymbol{a}_3 & \boldsymbol{a}_4 & \ldots & \boldsymbol{a}_n & \boldsymbol{a}_1 \end{pmatrix},$$

by abbreviation $\mathbf{A}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$, where $\boldsymbol{a}_i = [\underline{a}_i, \overline{a}_i]$, $\underline{a}_i \leq \overline{a}_i$ for each $i \in N$ we define *the interval circulant matrix* $\mathbf{A}^C(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ as the set of all circulant matrices from $\mathbf{A}$, in notation $\mathbf{A}^C = \{A \in \mathbf{A}; A \text{ is circulant}\}$.

Notice that there exist matrices $A \in \mathbf{A}$ which are not circulant. Since $\underline{A}$, $\overline{A} \in \mathbf{A}^C$, the set $\mathbf{A}^C$ is non-empty.

Definition 5. Let d be a positive natural number. An interval circulant matrix $\mathbf{A}^C(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ is called

- *possibly d-periodic* if there exists a matrix $A \in \mathbf{A}^C$ such that $\operatorname{per} A | d$,
- *universally d-periodic* if for each matrix $A \in \mathbf{A}^C$ $\operatorname{per} A | d$ holds.

3.1 Possible d-periodicity of interval circulant matrices

Let us denote $\tilde{S} = \{\overline{a}_i; \max_{k\in N} \underline{a}_k \leq \overline{a}_i\}$. It is clear that $\max_{k\in N} \overline{a}_k \in \tilde{S}$, so $\tilde{S} \neq \emptyset$. Let us define the number $\tilde{h} = \min \tilde{S}$ and the vector $\tilde{a} = (\tilde{a}_i) \in B(n)$ as follows:

$$\tilde{a}_i = \min\{\tilde{h}, \overline{a}_i\} \tag{2}$$

for each $i \in N$. For a given vector $a \in \boldsymbol{a}$ let us denote $h^{(a)} = \max_{i\in N} a_i$ and $J(a) = \{i \in N; a_i = h^{(a)}\}$. It is easy to see that $h^{(\tilde{a})} = \tilde{h}$ and $J(\tilde{a}) = \{i \in N; \overline{a}_i \geq \tilde{h}\}$.

The following lemma creates the base for the proof of the necessary and sufficient condition formulated in Theorem 4.

Lemma 4. Let $a \in \boldsymbol{a}$ be arbitrary and $\tilde{a}$ be given by (2). Then $J(a) \subseteq J(\tilde{a})$.

Proof. If $J(\tilde{a}) = N$, then $J(a) \subseteq J(\tilde{a})$ trivially holds. If $J(\tilde{a}) \neq N$ then it follows from (2) that for each $i \notin J(\tilde{a})$ the inequality $\overline{a}_i < \tilde{h}$ hods true. Consequently $\max_{i\notin J(\tilde{a})} \overline{a}_i \notin \tilde{S}$, i.e., $\max_{k\in N} \underline{a}_k > \max_{i\notin J(\tilde{a})} \overline{a}_i$. Let $a \in \boldsymbol{a}$, $r \in J(a)$ be arbitrary. We get

$$a_r = \max_{k\in N} a_k \geq \max_{k\in N} \underline{a}_k > \max_{i\notin J(\tilde{a})} \overline{a}_i \geq \max_{i\notin J(\tilde{a})} a_i$$

which implies $r \in J(\tilde{a})$. Consequently $J(a) \subseteq J(\tilde{a})$ for each $a \in \boldsymbol{a}$. □

The above constructed vector $\tilde{a} = (\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_n)$ defines a matrix $\tilde{A}(\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_n) \in \mathbf{A}^C$, which plays a crucial role in checking the possible d-periodicity of the interval circulant matrix $\mathbf{A}^C$.

Theorem 4. *An interval circulant matrix* $\mathbf{A}^C(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ *is possibly* d*-periodic if and only if* $\operatorname{per}\tilde{A} \mid d$.

Proof. If $\tilde{A}$ is not d-periodic then $\operatorname{per}\tilde{\mathcal{K}}^1 \nmid d$ where $\tilde{\mathcal{K}}^1 \in \mathrm{SCC}^\star(G(\tilde{A}, \tilde{h}))$ containing node 1. Let $A(a_1, a_2, \ldots, a_n)$ be arbitrary. It follows from Lemma 4 that $G(A, h^{(a)}) \subseteq G(\tilde{A}, \tilde{h})$. Then $\operatorname{per}\tilde{\mathcal{K}}^1 \mid \operatorname{per}\mathcal{K}^1$ where $\mathcal{K}^1 \in \mathrm{SCC}^\star(G(A, h^{(a)}))$ which implies $\operatorname{per}\mathcal{K}^1 \nmid d$. By Theorem 3 the matrix A is not d-periodic. This means that no matrix $A \in \mathbf{A}^C$ is d-periodic. Thus $\mathbf{A}^C$ is not possibly d-periodic.

The converse implication is trivial. □

The theorem bellow was proved in [3]. We will use the formula to compute the period of a circulant matrix in our algorithms. Denote $I(A) = \{i; a_{i+1} = \max_{k\in N} a_k\} \cup \{n\}$.

Theorem 5. [3] *Let* $A(a_0, \ldots, a_{n-1})$ *be a circulant matrix, let* $I(A) = \{n, i_0, i_1, \ldots, i_{k-1}\}$, $|I(A)| = k+1$. *Then*

$$\operatorname{per} A = \gcd\left(\frac{n}{\gcd(n, i_0)}, \frac{i_0 - i_1}{\gcd(i_0, i_1)}, \frac{i_1 - i_2}{\gcd(i_1, i_2)}, \ldots, \frac{i_{k-1} - i_{k-2}}{\gcd(i_{k-2}, i_{k-1})}\right). \tag{3}$$

Now, we can describe an algorithm based on Theorem 4 for checking the possible d-periodicity of an interval circulant matrix $\mathbf{A}^C$.

Algorithm **Possible d-periodicity of circulant matrix**

Input. $\mathbf{A} = [\underline{A}, \overline{A}]$ and d.
Output. 'yes' in variable *pdp* if $\mathbf{A}^C$ is possibly d-periodic; 'no' in *pdp* otherwise.
begin

(i) Compute $\tilde{S} = \{\overline{a}_i; \max_{k\in N} \underline{a}_k \leq \overline{a}_i\}$;
(ii) Compute $\tilde{h} = \min \tilde{S}$;
(iii) Compute the matrix $\tilde{A}$ by $\tilde{a}_i = \min\{\tilde{h}, \overline{a}_i\}$;
(iv) Compute the period $\operatorname{per}\tilde{A}$ (by (3));
(v) **If** $\operatorname{per}\tilde{A} \mid d$ **then** *pdp* :='yes'; **else** *pdp* :='no';

end

Theorem 6. *Let* $\mathbf{A}^C(\mathbf{a}_1,\ldots,\mathbf{a}_n)$ *be an interval circulant matrix. The Algorithm* **Possible** d**-periodicity of circulant matrix** *correctly decides in* $O(n)$ *time whether the interval circulant matrix* $\mathbf{A}^C$ *is possibly* d*-periodic.*

Proof. It was proved in [3] that the period of a circulant matrix can be computed in $O(n)$ time. Since none of operations in the Algorithm requires more time the total time of computation is $O(n)$. □

Let us consider the following interval matrix to illustrate the above Algorithm.

Example 1. Let $d = 4$. Let $\mathbf{A}([0,2],[2,5],[3,3],3,4],[1,3],[5,6])$ be an interval matrix.

$\tilde{S} = \{\overline{a}_i; \max_{k\in N} \underline{a}_k \le \overline{a}_i\} = \{5,6\}$ and $\tilde{h} = \min \tilde{S} = 5$. Now, the matrix $\tilde{A}$ can be found by $\tilde{a}_i = \min\{\tilde{h}, \overline{a}_i\}$. Hence the resulting circulant matrix is $\tilde{A}(2,5,3,4,3,5)$. $I(A) = \{i; a_i = \max_{k\in N} a_k\} \cup \{n\} = \{1,5,6\}$ and now, the period $\operatorname{per}\tilde{A}$ can be computed by (3):

$$\operatorname{per}\tilde{A} = \gcd\left(\frac{6}{\gcd(6,1)}, \frac{-4}{\gcd(1,5)}\right) = \gcd(6,-4) = 2.$$

Since $\operatorname{per}\tilde{A} \mid d$ the interval circulant matrix $\mathbf{A}^C$ is possibly d-periodic.

3.2 Universal d-periodicity of interval circulant matrices

The necessary and sufficient condition for universal d-periodicity of an interval circulant matrix is formulated in the following theorem.

Theorem 7. *An interval circulant matrix* $\mathbf{A}^C(\mathbf{a}_1,\ldots,\mathbf{a}_n)$ *is universally* d*-periodic if and only if* $\operatorname{per}\underline{A} \mid d$ *and* $(\forall k)(\overline{a}_k > \max_{i\in N} \underline{a}_i \Rightarrow \ell(c_k^1) \mid d)$*, where* $c_k^1 \in C^{(k)}$ *containing node* 1*.*

Proof. Suppose that $\operatorname{per}\underline{A} \nmid d$ or there exists $k \in N$ such that $\overline{a}_k > \max_{i\in N} \underline{a}_i$ and $\ell(c_k^1) \nmid d$. We shall prove that there exists $A \in \mathbf{A}^C$ such that $\operatorname{per} A \nmid d$.

Since $\underline{A} \in \mathbf{A}^C$, in the first case there exists $A \in \mathbf{A}$ such that $\operatorname{per} A \nmid d$.

In the second case we construct the matrix $\tilde{A} = (\tilde{a}_i)$ as follows:

$$\tilde{a}_i = \begin{cases} \overline{a}_i, & \text{if } i = k, \\ \underline{a}_i, & \text{otherwise,} \end{cases}$$

where $k \in N$ is such that $\overline{a}_k > \max_{i\in N} \underline{a}_i$ and $\ell(c_k^1) \nmid d$. Since $\mathcal{K}^1 \in \mathrm{SCC}^\star(G(\tilde{A}, \overline{a}_k))$ containing node 1, consists of only cycle c_k^1 we have $\operatorname{per}\mathcal{K}^1 = \ell(c_k^1) \nmid d$. Thus $\operatorname{per}\tilde{A} \nmid d$ by Theorem 3.

For the converse implication suppose that there exists $A \in \mathbf{A}^C$ such that $\operatorname{per} A \nmid d$ and $\operatorname{per}\underline{A} \mid d$. By Theorem 3 we get $\operatorname{per}\mathcal{K}^1 \nmid d$, where $\mathcal{K}^1 \in \mathrm{SCC}^\star(G(A, \max_{i\in N} a_i))$. As $\operatorname{per}\underline{A} \mid d$, we get $\max_{i\in N} a_i > \max_{i\in N} \underline{a}_i$. Let $k \in N$ be such that $a_k = \max_{i\in N} a_i$. Since A is circulant, $c_k^1 \in \mathrm{SCC}^\star(G(A, \max_{i\in N} a_i))$. Thus there exists $k \in N$ such that $\overline{a}_k \ge a_k > \max_{i\in N} \underline{a}_i$ and $\ell(c_k) \nmid d$. □

Now, we can describe an algorithm based on Theorem 7 for checking the universal d-periodicity of an interval circulant matrix $\mathbf{A}^C$.

Algorithm **Universal d-periodicity of circulant matrix**

Input. $\mathbf{A} = [\underline{A}, \overline{A}]$ and d.
Output. 'yes' in variable *udp* if $\mathbf{A}^C$ is universally d-periodic; 'no' in *udp* otherwise.

begin

(i) $k := 0$;

(ii) Compute per $\underline{A}$ (by (3));
(iii) **If** per $\underline{A} \nmid d$ **then** *udp* :='no'; **go to** end;
(iv) Compute $\underline{a} = \max_{i \in N} \underline{a}_i$;
(v) $k := k + 1$;
(vi) **If** $k > n$ **then** *udp* :='yes'; **go to** end;
(vii) **If** $\overline{a}_k \leq \underline{a}$ **go to** (v);
(viii) Compute $l(c_k)$ (by (1));
(ix) **If** $l(c_k) \nmid d$ **then** *udp* :='no'; **go to** end; **else go to** (v);
end

Theorem 8. *Let* $\mathbf{A}^C(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ *be an interval circulant matrix. The Algorithm* **Universal d-periodicity of circulant matrix** *correctly decides in* $O(n \log n)$ *time whether the interval circulant matrix* $\mathbf{A}^C$ *is universally d-periodic.*

Proof. The computational complexity of the period of a circulant matrix by (3) is $O(n)$ ([3]). Therefore the total time in steps (i)-(iv) is $O(n)$. To evaluate the length of a cycle by (1) requires $O(\log n)$ operations and this will be repeated n-times. Hence the complexity of the complete algorithm is $O(n \log n)$. □

Let us consider the following interval matrix to illustrate the above Algorithm.

Example 2. Let $d = 4$. Let $\mathbf{A}([0,4],[1,3],[4,6],[0,2],[2,3],[3,5],[4,4],[2,4])$ be an interval matrix.

For the circulant matrix $\underline{A} = (0,1,4,0,2,3,4,2)$ we find by (3) the period $\operatorname{per} \underline{A} = 2$. Since $\operatorname{per} \underline{A} \mid d$ we proceed to the next step and compute $\underline{a} = \max_{i \in N} \underline{a}_i = 4$. There are only two indices satisfying $\overline{a}_k > \underline{a}$ for which the computation of cycle length $l(c_k)$ is needed. For $k = 3$ is $l(c_k) = 4$ hence $l(c_k) \mid d$, while for $k = 6$ is $l(c_k) = 8$ and $l(c_k) \nmid d$. Thus there is a matrix $A \in \mathbf{A}^C$ with $\operatorname{per} A$ equal to 8. Consequently the considered interval circulant matrix $\mathbf{A}^C$ is not universally d-periodic, for $d = 4$.

A slightly modified interval matrix ($\boldsymbol{a}_5$ and $\boldsymbol{a}_6$ replace each other) of the matrix in previous example results in an universally d-periodic interval circulant matrix.

Example 3. Let $d = 4$. Let $\mathbf{A}([0,4],[1,3],[4,6],[0,2],[3,5],[2,3],[4,4],[2,4])$ be an interval matrix.

Instead of index $k = 6$ we shall consider index $k = 5$ for which the inequality $\overline{a}_k > \underline{a}$ is satisfied. Since $l(c_k) = 2$ divides $d = 4$, the given interval circulant matrix is universally d-periodic, for $d = 4$.

4 Conclusion

Polynomial algorithm for checking the possible d-periodicity with essentially improved computational complexity compared with interval matrices in general ([5]) was presented. Moreover, another polynomial algorithm for verifying the universal d-periodicity of interval fuzzy matrices was described while the computational complexity of corresponding procedure can be exponentially large in general.

References

[1] Fiedler, M., Nedoma, J., Ramík, J., Rohn, J. and Zimmerman, K.: *Optimization Problems with Inexact Data*. Springer–Verlag, Berlin, 2006.

[2] Garfinkel, R.: Minimizing wallpaper waste, *Oper. Res.* **25** (1977), 741-751.

[3] Gavalec, M.: *Periodicity in extremal algebras*. GAUDEAMUS, Hradec Králové, 2004.

[4] Molnárová, M., Myšková, H. and Plavka, J.: The robustness of interval fuzzy matrices, (DAM, article in press).

[5] Molnárová, M., Myšková, H. and Plavka, J.: On the periodicity of interval fuzzy matrices, MME, Karviná (2012).

[6] Plavka, J. and Szabó, P.: On the λ-robustness of matrices over fuzzy algebra, *DAM*, Volume **159** Issue **5** (2011), 381–388.

Periodicity of interval matrices in fuzzy algebra

M. Molnárová [1], H. Myšková [2], J. Plavka[3]

Abstract. The fuzzy algebra $\mathcal{B}$ is the triple $(B, \oplus, \otimes)$, where $(B, \leq)$ is a bounded linearly ordered set with binary operations *maximum* and *minimum*. The complete describing of d-periodic interval matrices over fuzzy algebra (fuzzy matrices) is presented and d-periodicity properties are proved. Characterization of the d-periodicity of interval fuzzy matrices is described and an $O(n^5)$ algorithm for checking the possible d-periodicity of interval fuzzy matrices is suggested.

Keywords: fuzzy algebra, periodicity, interval matrix, algorithm

JEL classification: C04
AMS classification: 08A72, 90B35, 90C47

1 Background of the problem

Fuzzy discrete dynamic systems can be introduced by fuzzy matrices and are useful for describing diagnosis of technical devices [9], medical diagnosis [8] or fuzzy logic programs [3]. The aim of this paper is to describe so called d-periodicity of matrices with inexact data (interval matrices) and to find algorithms for verifying the corresponding properties of interval matrices.
The fuzzy algebra $\mathcal{B}$ is the 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 and by $\mathbb{N}_0$ the set $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. The greatest common divisor of a set $S \subseteq \mathbb{N}$ is denoted by $\gcd S$. 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 $n \in \mathbb{N}$, $B(n, n)$ denotes the set of all square matrices of order 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 is denoted by A^r, with elements $a_{ij}^{(r)}$.

For A, $C \in B(n, n)$ we write $A \leq C$ ($A < C$) if $a_{ij} \leq c_{ij}$ ($a_{ij} < c_{ij}$) holds for all i, $j \in N$.

For a matrix $A \in B(n, n)$ the symbol $G(A) = (N, E_G)$ stands for a complete, arc-weighted digraph associated with A, i.e., the node set of $G(A)$ is N, and the capacity of any arc (i, j) is a_{ij}. Let $\emptyset \neq \tilde{N} \subset N$. $G/\tilde{N}$ stands for the subdigraph of digraph $G(A) = (N, E_G)$ with the node set $\tilde{N}$ and arc set $E_{G/\tilde{N}} = \{(i, j) \in E_G;\ i, j \in \tilde{N}\}$. A path in the digraph $G(A) = (N, E_G)$ is a sequence of nodes $p = (i_1, \ldots, i_{k+1})$ such that $(i_j, i_{j+1}) \in E_G$ 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. A digraph $G = (N, E_G)$ without cycles is called *acyclic*. If $G = (N, E_G)$ contains at least one cycle G is called *cyclic*.

By a *strongly connected component* $\mathcal{K}$ of $G(A, h) = (N, E_G)$ we mean a subdigraph $\mathcal{K}$ generated by a non-empty subset $K \subseteq N$ such that any two distinct nodes $i, j \in K$ are contained in a common cycle and K is a 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

[1]Technical University, Department of Mathematics and Theoretical Informatics, Němcovej 32, Košice, Slovakia, e-mail: Monika.Molnarova@tuke.sk
[2]Technical University, Department of Mathematics and Theoretical Informatics, Němcovej 32, Košice, Slovakia, e-mail: Helena.Myskova@tuke.sk
[3]Technical University, Department of Mathematics and Theoretical Informatics, Němcovej 32, Košice, Slovakia, e-mail: Jan.Plavka@tuke.sk

$\mathcal{K}$ 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. The set of all strongly connected components of G is denoted by $\mathrm{SCC}(G)$.

Definition 1. Let $A \in B(n,n)$. The *matrix period*, in notation: $\operatorname{per} A$ is defined as the minimal natural number p for which there is $R \in \mathbb{N}$ such that

$$A^{k+p} = A^k \quad \text{for all} \quad k \geq R.$$

By linearity of B, any element of any power of the matrix A is equal to some element of A. Therefore, the sequence of powers of A contains only finitely many different matrices with entries of A. As a consequence, the period of A is always well defined.

For given $h \in B$, the *threshold digraph* $G(A,h)$ is the digraph with the node set N and with the arc sets $E_G = \{(i,j);\ i,j \in N,\ a_{ij} \geq \mathrm{h}\}$.

Theorem 1. [4] *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 $G(A,h_2) \subseteq G(A,h_1)$.

Theorem 2. [2] *Let $A \in B(n,n)$, $d \in \mathbb{N}$. Then*

i) $\operatorname{per} A | d \Leftrightarrow (\forall h \in B)(\forall \mathcal{K} \in \mathrm{SCC}^\star(G(A,h)))\ \operatorname{per}\mathcal{K} | d$,

ii) $\operatorname{per} A = \operatorname{lcm}\{\operatorname{per}\mathcal{K}; \mathcal{K} \in \mathrm{SCC}^\star(A)\}$.

2 Periodicity of interval matrices

In this section we shall deal with matrices with interval elements. Similarly to [6], [7] we define an interval matrix $\mathbf{A}$.

Definition 2. Let $\underline{A}, \overline{A} \in B(n,n)$, $\underline{A} \leq \overline{A}$. An interval matrix $\mathbf{A}$ with bounds $\underline{A}$ and $\overline{A}$ is defined as follows

$$\mathbf{A} = [\underline{A}, \overline{A}] = \left\{\, A \in B(n,n);\ \underline{A} \leq A \leq \overline{A} \,\right\}.$$

Definition 3. An interval matrix $\mathbf{A}$ is called

- *possibly d-periodic* if there exists matrix $A \in \mathbf{A}$ such that $\operatorname{per} A | d$,
- *universally d-periodic* if for each matrix $A \in \mathbf{A}$ $\operatorname{per} A | d$ holds.

2.1 Possible d-periodicity

In this part we will prove a sufficient and necessary condition for an interval matrix to be possibly d-periodic. In addition we introduce a polynomial algorithm to check the possible d-periodicity and find the matrix $A \in \mathbf{A}$ such that $\operatorname{per} A | d$ in positive case.

Theorem 3. *Let $G' \subseteq G$, $\mathcal{K} \in \mathrm{SCC}^\star(G)$ and $\mathcal{K}' \in \mathrm{SCC}^\star(G'/N_{\mathcal{K}})$. Then* $\operatorname{per}\mathcal{K} | \operatorname{per}\mathcal{K}'$.

Proof. Since $\{\,\ell(c);\ c \text{ is a cycle from } \mathcal{K}'\,\} \subseteq \{\,\ell(\tilde{c});\ \tilde{c} \text{ is a cycle from } \mathcal{K}\,\}$ we have
$\operatorname{per}\mathcal{K} = \gcd\{\,\ell(c);\ c \text{ is a cycle from } \mathcal{K}\,\} | \gcd\{\,\ell(\tilde{c});\ \tilde{c} \text{ is a cycle from } \mathcal{K}'\,\} = \operatorname{per}\mathcal{K}'$. □

Denote $H = \{\overline{a}_{ij}; i,j \in N\} = \{h^{(1)}, h^{(2)}, \ldots, h^{(r)}\}$ where $h^{(1)} > h^{(2)} > \cdots > h^{(r)}$.

Theorem 4. *An interval matrix $\mathbf{A}$ is possibly d-periodic if and only if for each $h \in H$ and for each $\mathcal{K} \in \mathrm{SCC}^\star(G(\overline{A},h))$ such that $\operatorname{per}\mathcal{K} \nmid d$ the digraph $G(\underline{A},h)/N_{\mathcal{K}}$ is acyclic.*

Proof. Suppose that there exist $h \in H$ and $\mathcal{K} \in \mathrm{SCC}^{\star}(G(\overline{A}, h))$ such that $\mathrm{per}\,\mathcal{K} \nmid d$ and the digraph $G(\underline{A}, h)/N_{\mathcal{K}} \subseteq \mathcal{K}$ contains a cycle c. Let $A \in \mathbf{A}$ be arbitrary but fixed. As $G(\underline{A}, h) \subseteq G(A, h)$, there exists $\mathcal{K}^{'} \in \mathrm{SCC}^{\star}(G(A, h))$ such that $c \in \mathcal{K}^{'}$.

Since $G(A, h) \subseteq G(\overline{A}, h)$ and $\mathcal{K}' \in \mathrm{SCC}^{\star}(G(A, h)/N_{\mathcal{K}})$ by Theorem 2 we get $\mathrm{per}\,\mathcal{K} \mid \mathrm{per}\,\mathcal{K}'$ which implies $\mathrm{per}\,\mathcal{K}' \nmid d$. By Theorem 2 we get $\mathrm{per}\,A \nmid d$. Consequently the interval matrix $\mathbf{A}$ is not possibly d-periodic.

For the converse implication suppose that the digraph $G(\underline{A}, h)/N_{\mathcal{K}}$ is acyclic for each $h \in H$ and for each $\mathcal{K} \in \mathrm{SCC}^{\star}(G(\overline{A}, h))$ such that $\mathrm{per}\,\mathcal{K} \nmid d$. We shall construct a matrix $A^* \in \mathbf{A}$ such that $\mathrm{per}\,A^* \mid d$. First, wel construct an axillary sequence of matrices $\{A^{(k)}\}_{k=0}^{r} = \{(a_{ij}^{(k)})\}_{k=0}^{r}$ recurrently as follows:

$$a_{ij}^{(0)} = \underline{a}_{ij} \text{ for each } i, j \in N, \tag{1}$$

$$a_{ij}^{(k+1)} = \begin{cases} h^{(k+1)} & \text{if } (i,j) \in \bigcup\limits_{s \in M} E_{\mathcal{K}^s} \text{ and } a_{ij}^{(k)} < h^{(k+1)}, \\ a_{ij}^{(k)} & \text{otherwise,} \end{cases} \tag{2}$$

for each $k \in \mathbb{N}_0$, where $\mathcal{K}^1, \mathcal{K}^2, \ldots, \mathcal{K}^m \in \mathrm{SCC}^{\star}(G(\overline{A}, h^{(k+1)}))$ are such that $\mathrm{per}\,\mathcal{K}^s | \, d$ for $s = 1, 2, \ldots, m$. To finish the proof, we need the following claim.

Claim. Let for each $h \in H$ and for each $\mathcal{K} \in \mathrm{SCC}^{\star}(G(\overline{A}, h))$ such that $\mathrm{per}\,\mathcal{K} \nmid d$ the digraph $G(\underline{A}, h)/N_{\mathcal{K}}$ be acyclic. Then for each $k, l \in \mathbb{N}$ such that $l \leq k \leq r$ holds $\mathrm{per}\,\mathcal{K}' \mid d$ for each $\mathcal{K}' \in \mathrm{SCC}^{\star}(G(A^{(k)}, h^{(l)}))$.

Proof of the Claim. By mathematical induction on k

(i) For $k = 1$ we prove that $\mathrm{per}\,\mathcal{K}' \mid d$ for each $\mathcal{K}' \in \mathrm{SCC}^{\star}(G(A^{(1)}, h^{(1)}))$.
Let us denote by $\mathcal{K}^1, \mathcal{K}^2, \ldots, \mathcal{K}^m$ the non-trivial strongly connected components of $G(\overline{A}, h^{(1)})$ such that $\mathrm{per}\,\mathcal{K}^s \mid d$, $s = 1, 2, \ldots, m$. By (2), $\mathcal{K}^s \in \mathrm{SCC}^{\star}(G(A^{(1)}, h^{(1)}))$ for $s \leq m$. Moreover $\mathrm{SCC}^{\star}(G(A^{(1)}, h^{(1)})) = \{\mathcal{K}^1, \mathcal{K}^2, \ldots, \mathcal{K}^m\}$. Consequently $\mathrm{per}\,\mathcal{K}' \mid d$ for each $\mathcal{K}' \in \mathrm{SCC}^{\star}(G(A^{(1)}, h^{(1)}))$.

(ii) Suppose that $\mathrm{per}\,\mathcal{K}' | \, d$ for each for each $\mathcal{K}' \in \mathrm{SCC}^{\star}(G(A^{(k)}, h^{(l)}))$, $l \leq k$. It is easy to see that the digraphs $G(A^{(k+1)}, h^{(l)})$ and $G(A^{(k)}, h^{(l)})$ are identical for each $l \in \mathbb{N}$, $l \leq k$. Consequently, $\mathrm{per}\,\mathcal{K}' \mid d$ for each $\mathcal{K}' \in \mathrm{SCC}^{\star}(G(A^{(k+1)}, h^{(l)}))$, $l \leq k$. It remains to prove that $\mathrm{per}\,\mathcal{K}' \mid d$ for each $\mathcal{K}' \in \mathrm{SCC}^{\star}(G(A^{(k+1)}, h^{(k+1)}))$. It follows from the fact that $\mathrm{SCC}^{\star}(G(A^{(k)}, h^{(k+1)})) = \{\mathcal{K} \in (G(\overline{A}, h^{(k+1)})); \mathrm{per}\,\mathcal{K} | d\}$. □

Now, we shall continue in the proof of the theorem. If we apply the assertion of the previous claim for $k = r$, we get $\mathrm{per}\,\mathcal{K} \mid d$ for each $\mathcal{K} \in \mathrm{SCC}^{\star}(G(A^{(r)}, h^{(l)}))$ and $l \leq r$. Let us set $A^* = A^{(r)}$. Since $G(\overline{A}, h^{(r)})$ is complete we get $a_{ij}^* \geq h^{(r)}$ for each $i, j \in N$. In order to prove that $\mathrm{per}\,A^* \mid d$ it remains to show that $\mathrm{per}\,\mathcal{K}' \mid d$ for each $\mathcal{K}' \in \mathrm{SCC}^{\star}(G(A^*, a_{ij}^*))$, for $i, j \in N$ such that $a_{ij}^* > h^{(r)}$, $a_{ij}^* \notin H$. It is clear that $G(A^*, a_{ij}^*) = G(A^*, h^{(p)})$, where $p \in \mathbb{N}$ is such that $h^{(p+1)} < a_{ij}^* < h^{(p)}$. Thus $\mathrm{per}\,\mathcal{K}' \mid d$ for each $\mathcal{K}' \in \mathrm{SCC}^{\star}(G(A^*, a_{ij}^*))$, $i, j \in N$. According to Theorem 2 we get $\mathrm{per}\,A^* \mid d$. Thus the interval matrix $\mathbf{A}$ is possibly d-periodic.

We can use the obtained results to derive an algorithm for checking the possible d-periodicity of a given interval matrix $\mathbf{A} = [\underline{A}, \overline{A}]$.

Algorithm **Possible d-periodicity**

Input. $\mathbf{A} = [\underline{A}, \overline{A}]$.
Output. 'yes' in variable *pp* if $\mathbf{A}$ is possibly d-periodic; 'no' in *pp* otherwise.
begin

(i) Order the elements of H in such a way that $h^{(1)} > h^{(2)} > \cdots > h^{(r)}$;

(ii) $k := 0$; $A^{(k)} := \underline{A}$;

(iii) Find all strongly connected components of $G(\overline{A}, h^{(k+1)})$ and for each $\mathcal{K} \in \mathrm{SCC}^{\star}(G(\overline{A}, h^{(k+1)}))$ compute $\mathrm{per}\,\mathcal{K}$;

(iv) **If** there exists $\mathcal{K} \in SCC(G(\overline{A}, h^{(k+1)}))$ such that $\operatorname{per}\mathcal{K} \nmid d$ and the digraph $G(\underline{A}, h^{(k+1)})/N_{\mathcal{K}}$ is cyclic, **then** $pp :=' no'$; **go to** end;

(v) **Compute** $A^{(k+1)}$ (by (2));

(vi) $k := k + 1$;

(vii) **If** $k = r + 1$ **then** $pp :=' yes'$; $A^* := A^{(r)}$ **else go to** step (iii);

end

Theorem 5. *Let* $\mathbf{A}$ *be an interval matrix. The algorithm* **Possible** d**-periodicity** *correctly decides whether a matrix* $\mathbf{A}$ *is possibly* d*-periodic and finds* d*-periodic matrix* A^* *in* $O(n^5)$ *arithmetic operations.*

Proof. To determine the complexity of the algorithm, let us recall the well-known $O(n^2)$ algorithms to find all strongly connected components of a given digraph. If the strongly connected components are known we can compute the period of each of them using the $O(n^2)$ algorithm described by Balcer and Veinott in [1]. The number of operations for checking the strong connectivity and computing the periods for a given $h \in H$ is $O(n^2 r) \leq O(n^3)$. Thus, the complexity of the complete algorithm is $|H|O(n^3) \leq n^2 O(n^3) = O(n^5)$. □

Let us consider the following interval matrix to illustrate the Possible d-periodicity algorithm.

Example 2.1. Let $O = 0$, $I = 10$ and

$$\mathbf{A} = \begin{pmatrix} [0,0] & [3,4] & [0,0] & [0,5] & [0,0] \\ [2,2] & [0,0] & [2,3] & [0,0] & [0,0] \\ [2,3] & [0,0] & [0,2] & [0,0] & [0,0] \\ [0,0] & [0,0] & [0,0] & [0,1] & [1,2] \\ [0,0] & [0,0] & [1,1] & [2,2] & [0,0] \end{pmatrix}.$$

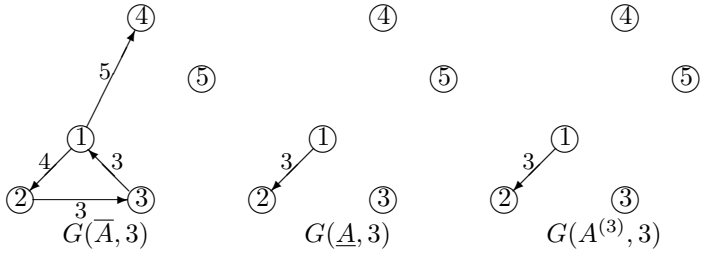


Figure 1: Threshold digraphs for $h^{(3)} = 3$

For $h^{(1)} = 5$ and $h^{(2)} = 4$ the digraphs $G(\overline{A}, h^{(1)})$, $G(\overline{A}, h^{(2)})$ are acyclic, so $A^{(1)} = A^{(2)} = \underline{A}$.
For $h^{(3)} = 3$ the digraphs $G(\overline{A}, h^{(3)})$ and $G(\underline{A}, h^{(3)})$ are presented on Figure 1.
We can see that $G(\overline{A}, h^{(3)})$ contains exactly one non-trivial strongly connected component with period $\operatorname{per}\mathcal{K} = 3 \nmid 4$. As $G(\underline{A}, 3)/N_{\mathcal{K}}$ is acyclic, the condition of Theorem 4 is satisfied and $A^{(3)} = \underline{A}$.
For $h^{(4)} = 2$ there are two strongly connected components in $G(\overline{A}, h^{(4)})$: $\mathcal{K}_1$ with $N_{\mathcal{K}_1} = \{1, 2, 3\}$, $\operatorname{per}\mathcal{K}_1 = 1$ and $\mathcal{K}_2$ with $N_{\mathcal{K}_2} = \{4, 5\}$, $\operatorname{per}\mathcal{K}_2 = 2$ (see Figure 2). As $\operatorname{per}\mathcal{K}_1 \mid 4$ and $\operatorname{per}\mathcal{K}_2 \mid 4$ we compute the matrix $A^{(4)}$ from $A^{(3)}$ by increasing elements $a_{33}^{(3)}$ and $a_{45}^{(3)}$ to 2. On Figure 3 we can see that for $h^{(5)} = 1$ the digraph $G(\overline{A}, 1)$ is strongly connected with period equal to one, so we compute the matrix $A^{(5)}$ from $A^{(4)}$ by increasing elements $a_{14}^{(4)}$ and $a_{44}^{(4)}$ to 1. We get

$$A^{(4)} = \begin{pmatrix} 0 & 3 & 0 & 0 & 0 \\ 2 & 0 & 2 & 0 & 0 \\ 2 & 0 & \mathbf{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{2} \\ 0 & 0 & 1 & 2 & 0 \end{pmatrix} \text{ and } A^{(5)} = \begin{pmatrix} 0 & 3 & 0 & \mathbf{1} & 0 \\ 2 & 0 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} & 2 \\ 0 & 0 & 1 & 2 & 0 \end{pmatrix}.$$

For $h^{(6)} = 0$ the digraph $G(\overline{A}, 0)$ is complete and $A^{(6)} = A^{(5)}$. Since $\operatorname{per} A^{(5)} = 2 \mid 4$, the matrix $A^{(5)}$ is d-periodic and consequently the given interval matrix $\mathbf{A}$ is possibly d-periodic.

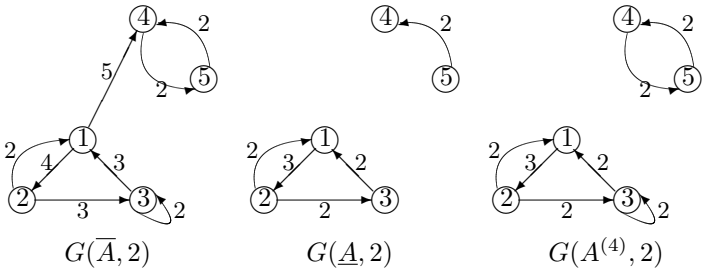


Figure 2: Threshold digraphs for $h^{(4)} = 2$

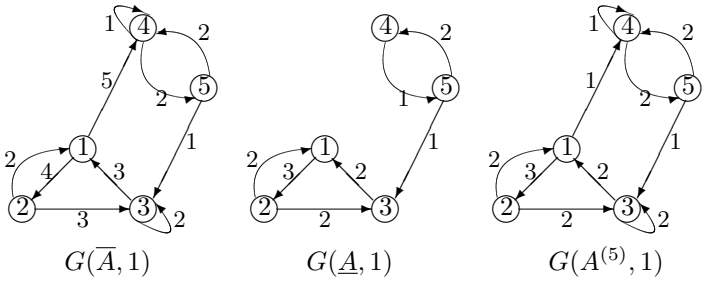


Figure 3: Threshold digraphs for $h^{(5)} = 1$

2.2 Universal d-periodicity

In this part we prove a necessary and sufficient condition for an interval matrix to be universally d-periodic.

For a given $h \in H$ let us denote $N^h = N \setminus \bigcup_{j=1}^{s} N_{\mathcal{K}^j}$, if $\mathrm{SCC}^\star(G(\underline{A}, h)) = \{\mathcal{K}^1, \ldots, \mathcal{K}^s\}$.

Theorem 6. *Let* $\mathbf{A}$ *be an interval matrix. Then* $\mathbf{A}$ *is universally* d*-periodic if and only if*

$$\operatorname{per} \underline{A} \mid d \quad \textit{and} \quad (\forall h \in H)(\forall c \in G(\overline{A}, h)/N^h)\,[\ell(c) \mid d].$$

Proof. Suppose that $\operatorname{per} \underline{A} \nmid d$ or there exist $h \in H$ and $c \in G(\overline{A}, h)/N^h$ such that $\ell(c) \nmid d$.
If $\operatorname{per} \underline{A} \nmid d$ then $\mathbf{A}$ is not universally d-periodic.
In the second case we construct the matrix $\tilde{A} = (\tilde{a}_{ij})$ as follows:

$$\tilde{a}_{ij} = \begin{cases} \overline{a}_{ij}, & \text{if } \ (i,j) \in c, \\ \underline{a}_{ij}, & \text{otherwise.} \end{cases}$$

There exists $\mathcal{K}^* \in \mathrm{SCC}^\star\, G(\tilde{A}, h)$ consisting of only cycle c. Since $\operatorname{per} \mathcal{K}^* = \ell(c) \nmid d$ by Theorem 2 we have $\operatorname{per} \tilde{A} \nmid d$. Thus an interval matrix $\mathbf{A}$ is not universally d-periodic.

For the converse implication we shall suppose that $\mathbf{A}$ is not universally d-periodic and $\operatorname{per} \underline{A} \mid d$. We prove that there exist $h \in H$ and $c \in G(\overline{A}, h)/N^h$ such that $\ell(c) \nmid d$.

If $\mathbf{A}$ is not universally d-periodic then there exist $A \in \mathbf{A}$, $h \in B$ and $\mathcal{K} \in \mathrm{SCC}^\star(G(A, h))$ such that $\operatorname{per} \mathcal{K} \nmid d$. Moreover $\operatorname{per} \underline{A} \mid d$ implies $N_{\mathcal{K}} \subseteq N^h$. From $\operatorname{per} \mathcal{K} \nmid d$ it follows that there exists a cycle $c \in \mathcal{K}$ such that $\ell(c) \nmid d$. From $G(A, h)/N^h \subseteq G(\overline{A}, h)/N^h$ it follows that $c \in G(\overline{A}, h)/N^h$. Define the value $\tilde{h}$ as follows:

$$\tilde{h} = \begin{cases} h^{(r)}, & \text{if } \ h \le \min\limits_{i,j \in N} \overline{a}_{ij} = h^{(r)}, \\ h^{(k)}, & \text{if } \ h^{(k)} \ge h > h^{(k+1)}. \end{cases}$$

Since $N^h \subseteq N^{\tilde{h}}$ and $G(\overline{A}, h) = G(\overline{A}, \tilde{h})$ we have $c \in G(\overline{A}, \tilde{h})/N^{\tilde{h}}$. Thus there exists a cycle $c \in G(\overline{A}, \tilde{h})/N^{\tilde{h}}$ such that $\ell(c) \nmid d$. □

Notice that Theorem 6 implies that the computational complexity of a procedure based on checking all cycles in $G(\overline{A}, h)/\tilde{N})$ can be exponentially large. The efficient algorithm for the interval circulant matrices is suggested in [5].

References

[1] Y. Balcer and A. F. Veinott, Computing a graph's period quadratically by node condensation, Discrete Math. 38 (1973) 295–303.

[2] M. Gavalec, Computing matrix period in max-min algebra, Discrete Appl. Math. 75 (1997) 63-70.

[3] T. Horvath and P. Vojtáš, Induction of fuzzy and annotated logic programs, Inductive Logic Programming 4455 (2007) 260-274.

[4] M. Molnárová, H. Myšková and J. Plavka, The robustness of interval fuzzy matrices (submitted to DAM).

[5] M. Molnárová, H. Myšková and J. Plavka, Efficient Algorithm for Checking Periodicity of Interval Circulant Fuzzy Matrices (MME 2012 Karviná)

[6] H. Myšková, Interval systems of max-separable linear equations, Linear Alg. Appl. 403 (2005), 263-272.

[7] H. Myšková, Control solvability of interval systems of max-separable linear equations, Linear Alg. Appl. 416 (2006), 215-223.

[8] E. Sanchez, Resolution of eigen fuzzy sets equations, Fuzzy Sets and Systems 1 (1978), 69-74.

[9] I. A. Zadeh: Toward a theory of fuzzy systems, In: R. E. Kalman, N. DeClaris, Eds., Aspects of Network and Systems Theory (Hold, Rinehart and Winston, New York, 1971), 209-245.

Using nonstationary time series for estimating small open economy model with financial frictions.

Tomáš Motl[1]

Abstract. This paper compares results of small open economy DSGE model estimation with prefiltered data and non-prefiltered data. There are at least two ways of taking a model to data: (i) filtering the historical time series outside the model in order to render them stationary or (ii) solving the model around balanced growth path and using nonstationary time series in estimation. While filtering is ubiquitous, there are number of problems associated with it. In particular, prefiltering time series outside of the model using univariate filters (the usual method) results in loss of information. This paper employs small open economy model with financial accelerator to show how prefiltering using univariate filters influences estimates of model parameters and the output gap. It concludes that for small-scale small open economy DSGE models, other ways of dealing with the filtering problem might be worth considering.

Keywords: DSGE, nonstationary, balanced growth path, filter

JEL classification: C51, E32
AMS classification: 91B66

1 Introduction

The dynamic stochastic general equilibrium models (DSGE) have become the standard tool for macroeconomic research and, increasingly, for policy evaluation and forecasting. However, the DSGE models are far from being undisputed or unchallenged. For general critique of DSGE models see Chari, Kehoe and McGrattan [7]. For assesment of the state-of-the-art models, see Schorfheide [14].

Criticisms of DSGE models often point out that the estimates of structural parameters are fragile. Schorfheide [12] compiled results from 42 studies that estimated DSGE model with Phillips curve on US data. He found that parameter estimates vary to great extent. Canova [5] focuses on the need to render the data stationary before estimating a DSGE model. He compares several univariate filtering techniques and finds that different techniques yield significantly different estimates of parameters. Furthermore, Canova shows that all investigated filters leave considerable spectral power outside the business cycle frequencies and points out some significant distortions that are likely to occur when filtering data.

Andrle [1] argues that it is useful to impose structural assumptions on the nature of trends in data, especially in developing economies, because permanent shocks have large influence on the business cycle and 'gap' models using detrended data are less likely to capture the true business cycle dynamics.

This paper compares the estimates of small open economy DSGE model on Czech data using Hodrick-Prescott filter and balanced growth path concept. Stationarizing and solving model around balanced growth path instead of fixed steady state is one possible remedy to the problem of detrending. I estimate two versions of small open economy DSGE model, one using data filtered by Hodrick-Prescott filter (HP model) and one using balanced growth path concept (BGP model) with nonstationary domestic technology. I show the difference in parameter estimates and esimated trajectory of output gap.

Section 2 provides short overview of the employed model. Section 3 describes data and shortly introduces the concept of balanced growth path. Section 4 shows and compares the results of estimation performed on nonstationary time series.

[1]Masaryk University, Department of Economics, Lipová 41a, Brno, tommotl@econ.muni.cz

2 The Model

I use DSGE model with financial accelerator as in Bernanke et al. [2]. The model is adjusted for open economy as in Shaari [13], while certain features are according to Justiniano and Preston [10]. The following section for the sake of space only introduces agents with their optimization problems. For details, consult Shaari [13].

Households Households maximize discounted expected utility given by

$$U(C_t, L_t) = \varepsilon_{G,t} \left(\log(C_t) - \frac{L_{H,t}^{1+\zeta}}{1+\zeta} \right),$$

where C_t is consumption, $L_{H,t}$ is labor supply, ζ is inverse elasticity of labor supply and $\varepsilon_{G,t}$ is preference shock.
Consumption is a bundle of domestic and foreign goods

$$C_t = \left[(1-\gamma)^{\frac{1}{\eta}} (C_{H,t})^{\frac{\eta-1}{\eta}} + \gamma \frac{1}{\eta} (C_{F,t})^{\frac{\eta-1}{\eta}} \right]^{\frac{\eta}{\eta-1}}$$

where γ measures preference for foreign goods (and openness of the domestic economy) and η measures elasticity of substitution between domestic goods (indexed by H) and foreign goods (indexed by F).
Household budget constraint is

$$\widetilde{W}_t L_t + R_{t-1} D_{t-1} + R_{t-1}^* \Psi^B(Z_{t-1}, \varepsilon_{t-1}^{UIP}) S_t B_{t-1} + \Pi_t = P_t C_t + D_t + S_t B_t$$

where D_t is one period domestic riskless bond that yields R_t, B_t is the foreign bond, S_t is the nominal exchange rate and Ψ^B is interest rate premium derived from net foreign asset position $Z_t = \frac{S_t B_t}{Y_t P_t}$ of the domestic economy. ε_t^{UIP} is UIP shock that like other shocks, with the exception of monetary policy shock, follows AR(1) process with parameter ρ_{UIP}.

Entrepreneurs Entrepreneurs manage intermediate firms, produce capital goods and own all capital. A fraction of $(1-\xi)$ of entrepreneurs die every period.
Intermediate firms produce intermediate goods using technology

$$Y_t = K_{t-1}^{\alpha} (L_t A_t)^{(1-\alpha)}, \tag{1}$$

where L_t is composite of households and entrepreneur labor

$$L_t = L_{H,t} \Omega L_{E,t}^{1-\Omega}$$

and we normalize $L_{E,t}$ to one for simplicity. A_t is labor-augumenting nonstationary technology.
Entrepreneurs produce capital using old capital and investment INV_t according to

$$K_{t+1} = \varepsilon_{I,t} INV_t + (1-\delta) K_t - \chi_I \left(\frac{INV_t}{K_t} - \delta \right)^2.$$

The holding of capital is financed partly by net worth N_t, partly by loans L_t:

$$Q_t K_{t+1} = F_{t+1} + N_{t+1}.$$

where Q_t is the real price of capital. When borrowing from finnacial intermediary, entrepreneurs pay premium according to their leverage. On the margin, they equate marginal revenue of additional unit of capital with marginal cost of financing that unit:

$$E_t \frac{\{R_{G,t+1} + (1-\delta) Q_{t+1}\} K_{t+1}}{Q_t K_{t+1}} = E_t \left[\left(\frac{N_{t+1}}{Q_t K_{t+1}} \right)^{-\psi} R_t \frac{P_t}{P_{t+1}} \right]$$

Retailers and Importers Importers buy foreign goods for price $P^W_{F,t} = S_t P^*_t$ and sell for price $P_{F,t} \neq S_t P^*_t$ in monopolistically competitive market. The law of one price does not hold. Only a fraction of $(1-\theta_F)$ of importers can change prices every period. The representative importer maximizes expected discounted value of future profits:

$$\max_{P^{NEW}_{F,t}} \sum_{k=0}^{\infty} (\beta\theta_F)^k E_t \left(Y_{F,t+k}(z) \left[P^{NEW}_{F,t} - P_{F,t+k} \frac{P^W_{F,t}}{P_{F,t+k}} \right] \right)$$

Retailers operate in similar fashion and face almost identical optimization problem with parameter θ_H. The overall domestic inflation is given by

$$\pi_t = (1-\gamma)\pi_{H,t} + \gamma\pi_{F,t}.$$

Monetary Policy, Fiscal Policy, Market Clearing Interest rate is given by

$$r_t = (1-\rho)\left[\psi_\pi \pi_{t+1} + \psi_y y_{t+1}\right] + \rho r_{t-1} + \varepsilon^{MP}_t.$$

Government pursues Ricardian fiscal policy with balanced budget. Market clearing requires

$$Y_{H,t} = \left(\frac{P_{H,t}}{P_t}\right)^{-\eta} \left[(1-\gamma)(C_t + INV_t) + \gamma \left(\frac{1}{RER_t}\right)^{-\eta} Y^*_t \right],$$

where RER_t is the real exchange rate.

Foreign Economy Foreign economy is assumed to be large closed version of the domestic economy. The main difference is that foreign intermediate firms produce goods using labor only, therefore there is no capital accumulation and financial frictions in the foreign economy. Simple specification of foreign economy does not influence the results and can be replaced by VAR(1) model. Variables and parameters in foreign economy are denoted by asterisk.

3 Balanced Growth Path, Data, Estimation

Balanced Growth Path The balanced growth path (henceforth BGP)is a trajectory along which all variables in the model grow at predefined (possibly different) constant rates. Nice summary of modeling BGP is provided by King, Plosser and Rebelo [11].

Selected necessary conditions for a DSGE model to satisfy BGP that are relevant for this paper are: (a) income effect equals substitution effect, which for usual utility functions means logarithmic utility in consumption; (b) labor-augumenting technological progress; (c) production function with constant returns to scale. Other necessary conditions are of low importance for common DSGE models (see King, Plosser and Rebelo [11]).

As an example, consider closed economy model with trends in technology (growth rate A) and money growth (growth rate M). Along BGP, variables grow with following growth rates: real quantities, marginal product of labor: A; nominal quantities: M; prices: $\frac{A}{M}$; hours, interest rates, marginal product of capital: stationary.

Before solving the model using standard methods, we need to stationarize it, that is, remove trends form the model. This means to divide each nonstationary variable by appropriate source of nonstationarity. Consider production function (1) and denote $a_t = \frac{A_t}{A_{t-1}}$, $k_t = \frac{K_t}{A_t}$ and $y_t = \frac{Y_t}{A_t}$. The equation (1) then becomes

$$y_t = \left(\frac{k_{t-1}}{a_t}\right)^{\alpha} (L_t)^{(1-\alpha)\Omega} \tag{2}$$

Applying this transformation to all equations yields stationary system that can be log-linearized and solved using standard methods. In this paper, I employ an approach that features one source of nonstationarity: domestic technology which growth rate fluctuates around steady state values g_A calibrated from data:

$$a_t = \rho_A a_{t-1} + (1-\rho_A) g_A + \varepsilon_{A,t}. \tag{3}$$

Data and estimation The model is estimated on seasonally adjusted Czech and eurozone historical time series (1Q2000-2Q2011) of GDP, inflation and interbank interest rate. For Czech economy I also include consumption and real exchange rate EUR/CKZ. For stationarized model, logarithmized time series are filtered by Hodrick-Prescott (HP) filter and expressed as percentage deviations from trend. For BGP model, time series of GDP and consumption are logarithmized. Other time series still need to be prefiltered. To avoid the contamination of results from BGP model by using HP filter and to highlight the differences, I detrend the time series by linear trend. Using both stationary and nonstationary time series is valid (see for example Iacoviello and Nero [9])

Both models are estimated using Matlab toolbox Dynare by generating 500,000 samples in each of 3 runs of Metropolis-Hastings algorithm. To allow for comparison, I set identical priors for all estimated parameters in both models. Convergence was checked using Brooks-Gelman diagnostics (see [3]) and prior-posterior plots. Both models were estimated succesfully.

4 Results

Table 1 provides summary of prior and posterior distributions of estimated parameters. It is obvious that parameter estimates are different. Analysis of impulse response functions (not included for sake of space) shows that difference in parameter estimates translates into difference in model behavior.

Estimates of preference parameters ζ and η show large differences, with uncertainty of estimates given by confidence intervals larger for HP model. Price stickiness parameters in HP model suggest higher price stickiness in domestic goods (θ_H) than in imported goods (θ_F), while BGP model suggests the opposite. Hloušek [8] studied similar question and found evidence in line with BGP model. While Taylor rule parameters for ψ_π and ψ_y are almost identical for foreign economy, they differ for the domestic economy. Here, again, BGP model estimates are closer to literature. Looking at the standard deviations of shocks, there are many differences in estimates. Most notably, HP model suggests similar standard errors for domestic and foreign monetary policy shock ($\sigma_{\varepsilon_{MP}}$), while BGP model suggests larger domestic shocks. This can be seen as evidence in favor of the HP model. Overall, both models provide plausible estimates of parameters with evidence from literature on both sides.

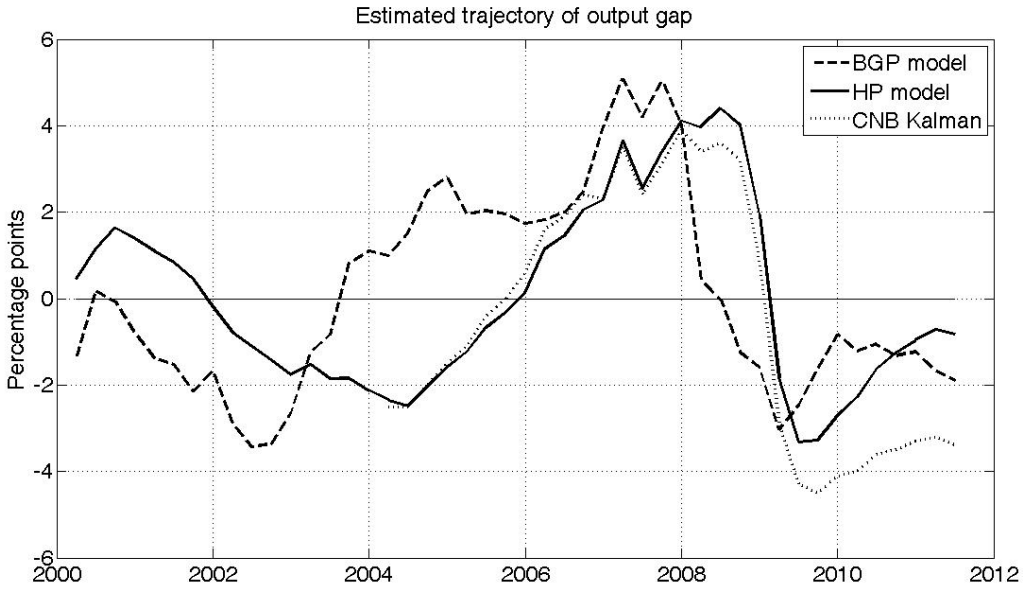


Figure 1 Various estimates of Czech output gap 1Q2000-2Q2011

Next, we can compare smoothed trajectory of output gap. I focus on the output gap because it is a variable of importance for monetary policy. At the same time, it is an unobserved variable and as such can be estimated by several methods. Figure 1 shows estimated output gaps from BGP and HP model. For comparison, I also include output gap estimated by the Czech National Bank (CNB) using Kalman

	Prior			BGP model			HP model		
Name	Distribution	Mean	Std	Mean	90% CI		Mean	90% CI	
ζ inv. labor elast.	gamma	2.00	0.50	0.562	0.351	0.789	1.137	0.619	1.629
η domestic/imports elast.	normal	0.50	0.10	0.363	0.305	0.423	0.155	0.105	0.206
θ_H domestic Calvo	beta	0.70	0.05	0.528	0.424	0.628	0.764	0.670	0.860
θ_F imports Calvo	beta	0.70	0.05	0.836	0.795	0.878	0.648	0.586	0.712
θ^* foreign Calvo	beta	0.70	0.05	0.799	0.755	0.844	0.792	0.751	0.832
χ_I investment adj. cost	gamma	12.00	1.50	12.261	9.789	14.710	8.618	6.692	10.507
ψ risk premium elast.	gamma	0.05	0.01	0.048	0.032	0.064	0.093	0.069	0.116
ψ_π Taylor rule - infl.	gamma	1.50	0.10	1.320	1.180	1.459	1.469	1.322	1.608
ψ_y Taylor rule - output	gamma	0.25	0.05	0.298	0.222	0.373	0.185	0.132	0.238
ψ^*_π foreign Taylor rule	gamma	1.50	0.10	1.422	1.258	1.578	1.416	1.259	1.571
ψ^*_y foreign Taylor rule	gamma	0.25	0.05	0.185	0.138	0.233	0.198	0.143	0.252
ρ int. rates smoothing	beta	0.70	0.05	0.701	0.619	0.782	0.890	0.866	0.915
ρ^* foreign smoothing	beta	0.70	0.05	0.817	0.778	0.851	0.814	0.779	0.852
ρ_G shock AR(1) param.	beta	0.50	0.05	0.569	0.484	0.657	0.597	0.524	0.667
ρ^*_G shock AR(1) param.	beta	0.50	0.15	0.752	0.661	0.841	0.633	0.529	0.740
ρ_A shock AR(1) param.	beta	0.50	0.15	0.343	0.234	0.451	0.593	0.378	0.812
ρ^*_A shock AR(1) param.	beta	0.50	0.15	0.533	0.331	0.737	0.390	0.199	0.578
ρ_I shock AR(1) param.	beta	0.50	0.15	0.882	0.832	0.932	0.500	0.386	0.618
ρ^*_{UIP} shock AR(1) param.	beta	0.50	0.15	0.635	0.533	0.743	0.853	0.793	0.916
$\sigma(\varepsilon_G)$ shock STD	inv. gamma	1	∞	3.348	2.688	4.000	1.171	0.943	1.388
$\sigma(\varepsilon^*_G)$ shock STD	inv. gamma	1	∞	1.749	1.435	2.060	1.567	1.290	1.848
$\sigma(\varepsilon_A)$ shock STD	inv. gamma	1	∞	0.960	0.688	1.228	7.022	1.656	13.342
$\sigma(\varepsilon^*_A)$ shock STD	inv. gamma	1	∞	8.895	2.357	16.284	1.6173	0.941	2.269
$\sigma(\varepsilon_I)$ shock STD	inv. gamma	1	∞	7.212	5.324	9.057	8.456	6.064	10.743
$\sigma(\varepsilon_{MP})$ shock STD	inv. gamma	0.1	∞	1.217	0.946	1.473	0.120	0.091	0.147
$\sigma(\varepsilon^*_{MP})$ shock STD	inv. gamma	0.1	∞	0.091	0.074	0.108	0.092	0.0744	0.108
$\sigma(\varepsilon_{UIP})$ shock STD	inv. gamma	1	∞	1.255	0.810	1.675	0.421	0.248	0.587

Table 1 Prior and posterior distributions of estimated parameters for BGP and HP model.

filter [1]. What is obvious is that BGP model provides trajectory that is very different from the other two. Output gap identified by Hodrick-Prescott filter is very similar to the one identified by Kalman filter at CNB.

Although the output gap is not observable and therefore there is no definite way of deciding which trajectory is the correct one, it is likely that the BGP model fares poorly relative to HP model. It is difficult to substantiate the zero output gap in middle of 2008, when inflation, while subsiding, was still running around 6 per cent and decline in foreign demand was still to come. Similarly, in the 2004-2006 period the BGP model suggests positive values of output gap, while both HP model and Kalman filter suggest negative to near-zero values. This discrepancy can be taken as evidence against the BGP model.

5 Conclusion

This paper briefly introduced the concept of balanced growth path in a small open economy DSGE model. The estimated parameters do not provide clear guidance as to whether using BGP concept is superior to using Hodrick-Prescott. However, the analysis of output gap trajectory suggests that filtering by Hodrick-Prescott filter does not introduce significant misspecifications. On the other hand, BGP model with one source of nonstationarity does not provide credible estimate of output gap.

It is likely that more sources of nonstationarity are needed for small open economy models of the Czech economy. It is plausible to consider nonstationary domestic and foreign technology and domestic and foreign price level. However, introducing four sources of nonstationarity brings along significant increase in computational complexity. The benefits of using BGP approach may indeed be greater in more complex, possibly nonlinear models.

Other possible ways eliminating influence of particular filter have been suggested. Brůha [4] specifies long-run trend relations to be estimated jointly with short-run fluctations. Canova and Ferroni [6] propose

[1] Source: CNB Inflation Report II/2012.

using multiple filters for estimation to relax the influence of each filtering technique. More observed time series are then linked to one model variable using weighted measurement errors. Such data treatment offers model-consistent way of eliminating influence of particular filtering method. In addition, it is also consistent with bayesian framework and data-rich environment and able to provide a way of working around some other issues with data (such as whether inflation should be measured by CPI or GDP deflator). This method is likely to deliver better results for small-scale small open economy model.

Acknowledgements

This work is supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/0780/2011.

References

[1] Andrle, M.: *The Role of Trends and Detrending in DSGE Models*. unpublished, available from: http://mpra.ub.uni-muenchen.de/13289/, 2008.

[2] Bernanke, B. S., Gertler, M. and Gilchrist, S.: The financial accelerator in a quantitative business cycle framework. In: *Handbook of Macroeconomics* (J.B. Taylor and M. Woodford, eds.), Elsevier, 1999, volume 1, chapter 21, 1341–1393.

[3] Brooks, S. P. and Gelman, A.: General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, (1998), 434–455.

[4] Brůha, J.: An Empirical Small Labor Market Model for the Czech Economy. *Czech Journal of Economics and Finance* **61(5)** (2011), 434–449.

[5] Canova, F.: *Bridging Cyclical DSGE Models and the Raw Data*. unpublished, available from: http://crei.cat/people/canova/dsge_trend(1).pdf, 2009.

[6] Canova, F., Ferroni, F.: Multiple filtering devices for the estimation of cyclical DSGE models. *Quantitative Economics* **2(1))** (2011), 73–98.

[7] Chari, V. V., Kehoe, P.J. and McGrattan, E. R.: New Keynesian models: not yet useful for policy analysis. *American Economic Journal: Macroeconomics*, **1(1)**, American Economic Association, (2009), 242–266.

[8] Hloušek, M.: Nominal rigidities and wage-price dynamics in DSGE model of open economy: application for the Czech Republic. *Doctoral Thesis*, Masaryk University, Brno, 2010.

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

[10] Justiniano, A. and Preston, B.: Monetary policy and uncertainty in an empirical small open-economy model. *Journal of Applied Econometrics*, John Wiley & Sons, Ltd., **25(1)** (2010), 93–128.

[11] King, R. G., Plosser, Ch. I. and Rebelo, S. T.: Production, Growth and Business Cycles: Technical Appendix. *Computational Economics* **20)** (2002), 87–116.

[12] Schorfheide, F.: DSGE model-based estimation of the new Keynesian Phillips curve. *Economic Quarterly* **94** (2008), 397–433.

[13] Shaari, M. H.: *Analyzing Bank Negara Malaysia's Behaviour in Formulating Monetary Policy: An Empirical Approach*. The Australian National University, Canberra, 2008.

[14] Schorfheide, F.: Estimation and Evaluation of DSGE Models: Progress and Challenges. *NBER Working Papers* **16781** (2011).

Multi-criteria evaluation of alternatives applied to the mobile phone tariffs in comparison with Monte-Carlo simulation results

Petr Mynařík[1], Martina Kuncová[2]

Abstract. In this article we try to compare the results of application the Monte Carlo simulation model and the multi-criteria evaluation of alternatives to the same example. The example contains data of different mobile phone tariffs. We will try to find out if it is possible to construct and solve the multi-criteria model with adjusted weights to find the same results as simulation model does. In our opinion this comparison can be very interesting, because multi-criteria evaluation of alternatives and simulation modeling (Monte Carlo simulation) are two different approaches of mathematical methods connected with the operational research. Monte Carlo simulation tries to iteratively evaluate the deterministic model by using random inputs. Methods of multi-criteria evaluation of alternatives use given inputs to find the order of the alternatives. This order is influenced mainly by weights of the criteria.

Keywords: multi-criteria evaluation of alternatives, Monte Carlo simulation, Mobile Phone Tariffs.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

In this article we will try to compare the application of Monte Carlo method and selected methods of multi-criteria evaluation of alternatives on the same example. Monte Carlo simulation and multicriteria evaluation are quite different approaches to finding a possible solution, but that's why we try to compare the subsequent analysis and compare the results of both approaches.

Then we will find out, if it is possible to use multi-criteria evaluation of alternative methods to obtain the same results as from the simulation model.

2 Methods

Before we start the analysis we have to select the alternatives (mobile operators' tariffs), the criteria and the distributions for the random variables generation. This analysis is focused on the specific situation – to find the best tariff for one employee of the Executive Board of the Czech Union for Nature Conservation to minimize the costs of telephone calls. The entire model for more employees has been created in the diploma thesis [4] where all (69 possible) the mobile operators' tariffs and their data are described.

There is problem in the case when we don't know preferences of user in any form. Also in such case one solution of this problem is a simulation of weights. We generate weight vectors, that would as best as possible describe the selected employees [4], [5]. These weight vectors will be used as input for multi-criteria evaluation methods. Then we can observe changing alternatives that are at the first places. As we would like to compare the results with the order of the tariffs created by the TOPSIS and PROMETHEE II.

2.1 Multiple criteria decision making

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. In this article I will use these methods: TOPSIS

[1] University of Economics, Department of Econometrics, W. Churchill Sq. 4, 130 67 Prague 3, petr.mynarik@vse.cz.
[2] University of Economics, Department of Econometrics, W. Churchill Sq. 4, 130 67 Prague 3, martina.kuncova@vse.cz.

and PROMETHEE [1], [2], [3]. Why did I choose these methods? Because I know very well these methods and I am sure that interesting results will be provided by them.

TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution):

The output provided by TOPSIS is a complete arrangement of possible alternatives with respect to the distance to both the ideal and the basal alternatives incorporating relative weights of criterion importance. The required input information includes decision matrix Y and weight vector v.

In addition, in the same way as in the WSA an assumption of maximization of all the criteria is true (otherwise it is necessary to make an appropriate transformation). This decision-making approach can be summarized in the following steps:

- normalize the decision matrix according to Euclidean metric:

$$r_{ij} = \frac{y_{ij}}{\sqrt{\left(\sum_{i=1}^{p}(y_{ij})^2\right)}}, \quad i = 1, 2, \dots, p, \quad j = 1, 2, \dots, k,$$

- calculate the weighted decision matrix W = (wij): $w_{ij} = v_j \cdot r_{ij}$,
- from the weighted decision matrix W identify vectors of the hypothetical ideal and basal alternatives over each criterion: $\mathbf{H} = (H_1, H_2, \dots, H_k)$ and $\mathbf{D} = (D_1, D_2, \dots, D_k)$, where $H_j = max_i\, w_{ij}$, $j = 1, 2, \dots, k$ and $D_j = min_i\, w_{ij}$, $j = 1, 2, \dots, k$,
- measure the Euclidean distance of every alternative to the ideal and to the basal alternatives over each attribute:

$$d_i^+ = \sqrt{\sum_{j=1}^{n}(w_{ij} - H_j)^2} \text{ and } d_i^- = \sqrt{\sum_{j=1}^{n}(w_{ij} - D_j)^2},$$

- for each alternative determine the relative ratio of its distance to the basal alternative:

$$c_i = \frac{d_i^-}{d_i^+ + d_i^-},$$

- rank order alternatives by maximizing ratio.

PROMETHEE (Preference Ranking Organization Method for Enrichment Evaluations):

PROMETHEE is a partial ranking of the actions. It is based on the positive and negative flows. It includes preferences, indifferences and incomparabilities (partial preorder).

The result of application of this method is the expression intensity of preferences between pairs of variants as measured by all criteria.

The method uses the preference function for expression of intensity of preferences.

- marginal range of functional values:.

  $P(a_i, a_j) = 0$ … indifference a_i and a_j

  $P(a_i, a_j) \sim 0$ … weak preferences a_i to a_j

  $P(a_i, a_j) \sim 1$ … large preferences a_i to a_j

  $P(a_i, a_j) = 1$ … absolute preferences a_i to a_j

- range of preference function depends on the difference of criteria values: **d** = f (a_i) – f(a_j)

  (greater difference = greater intensity of preference)

- there are defined some types of generalized criteria

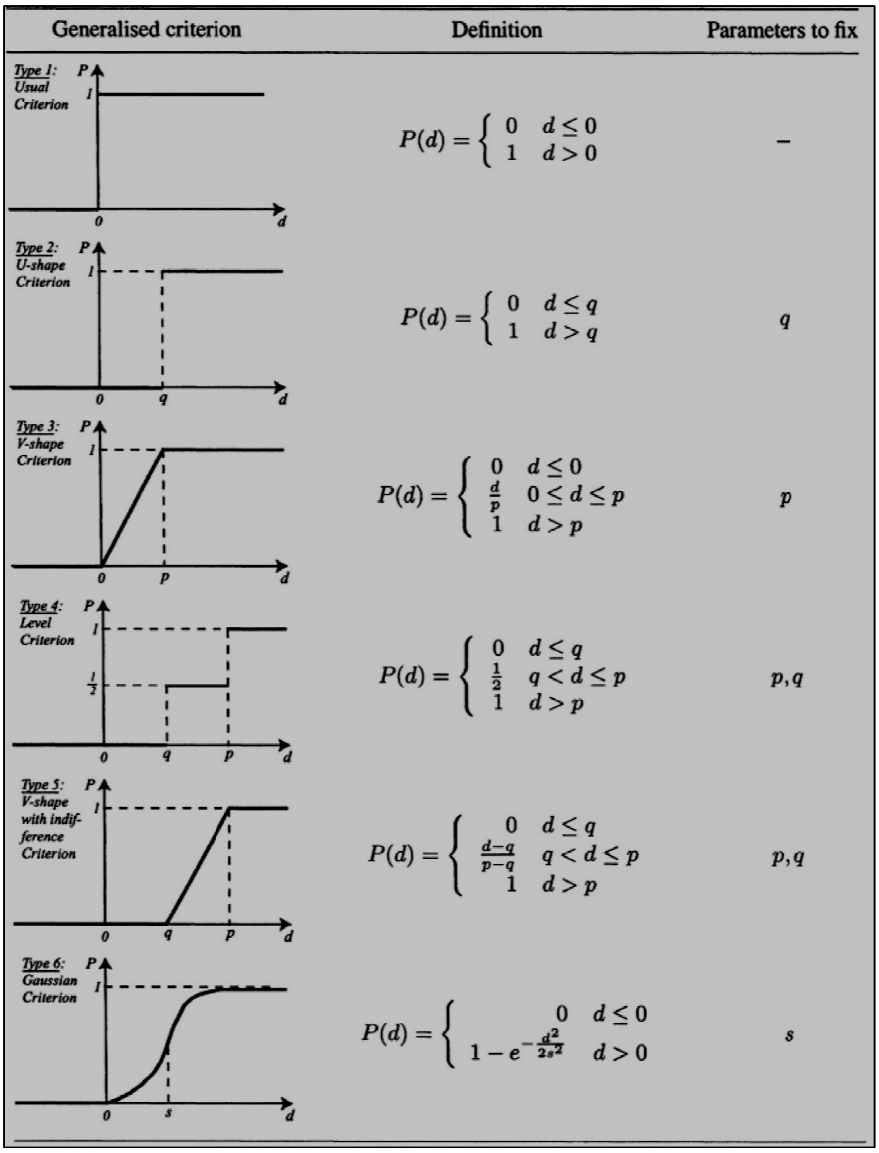


Figure 1 Types of generalized criteria (P(d): Preference function). [6]

2.2 Monte Carlo Simulation

Monte Carlo simulation (or technique) is closed to statistics as it is a repeated process of random sampling from the selected probability distributions that represent the real-life processes. On the basis of the existed information we should select the type of probability distribution that corresponds to our expectations and define all the parameters for.

The problem of some economic models is the lack of the information – especially in the retail sector sometimes only managers themselves know how the process works, what the typical number of customers during a period is etc. In this kind of situations we cannot use basic statistical or mathematical models as we do not have the strict or real data. That is why Monte Carlo simulation can help as it uses random variables from different distributions. This kind of simulation was used also in the diploma work [4] to find the best tariff. But it is possible to use it also to generate the weights of the criteria – or better to say generate the points for each criterion and then calculate the weights using the Point method [2].

3 Data - Mobile Operators' Tariffs

We used all the data and information from the diploma thesis [4]. We created 7 types of criteria to compare 69 tariffs. For Monte Carlo simulation all data in rough form can be used. However for multi-criteria decision-making we need to transform data in relevant form. We can enumerate all selected criteria:

1. fixed payment tariff – minimal (CZK)

2. the number of free minutes – maximal (minutes)

3. price for 1 minute calling in own net – minimal (CZK)

4. price for 1 minute calling in landline (phone) – minimal (CZK)

5. price for 1 minute calling in other net – minimal (CZK)

6. possibility of free calls to own network on weekdays – maximal (scale 0-10 points)

7. possibility of free calls to own network on weekends – maximal (scale 0-10 points)

Then we have founded out that from the 69 tariffs only 38 are non-dominated, so therefore we will compare only these non-dominated tariffs. But there is problem how to set up weights.

4 Analysis and results

The simulation model of the described situation showed [4], [5], that the best tariffs are from the O2 operator called "Podnikání L" and "Podnikání M". And we want to determine, whether the order of these two tariffs will be changeable after an application of different mathematical approach (multiple criteria decision making).

In this analysis we used software LINGO, where we set up a mathematical model a then we were searching particular weight vectors for different order of studied alternatives.

Next table shows results of application method TOPSIS a PROMETHEE II. In the analysis we were interested in only the order of these two variants - "Podnikání L" and "Podnikání M". You can see, that the weights are crucial for the results. We found out, that it is not possible to find a weight vector for the "Podnikání L" and "Podnikání M"tariff to be on the first place. The best position you can see on this table.

	Crit. 1	**Crit. 2**	**Crit. 3**	**Crit. 4**	**Crit. 5**	**Crit. 6**	**Crit. 7**	**Order - Podnikání L**		**Order - Podnikání M**	
								TOPSIS	**PROM. I**	**TOPSIS**	**PROM. I**
Weight vector 1	0.72	0.02	0.025	0.025	0.025	0.16	0.025	2.	3.	3.	5.
Weight vector 2	0.663	0.114	0.022	0.029	0.017	0.128	0.027	4.	3.	2.	1.
Weight vector 3	0.682	0.087	0.019	0.024	0.015	0.153	0.02	3.	4.	2.	3.

Table 1 The best order of the selected tariffs and the weight vectors

	Crit. 1	**Crit. 2**	**Crit. 3**	**Crit. 4**	**Crit. 5**	**Crit. 6**	**Crit. 7**
Type of generalized criteria	V-shape crit.	Usual crit.	Level crit.	Level crit.	Level crit.	V-shape crit.	V-shape crit.
Parameter	p = 1000		d = 1.5 p = 2.5	d = 1.5 p = 2.5	d = 1.5 p = 2.5	p = 3	p = 2

Table 2 Promethee II – preference function and parametres

Not only does the order of alternative depends on weight vector, but there is very important choice of preference function and value of parametres. This situation is described in the next table, where we changed types of generalized criteria and value of parametres. And you can see, that the order of alternatives is partially different.

	Crit. 1	Crit. 2	Crit. 3	Crit. 4	Crit. 5	Crit. 6	Crit. 7	Order – Podnikání L PROM. I	Order – Podnikání M PROM. I
Weight vector 1	0.72	0.02	0.025	0.025	0.025	0.16	0.025	3.	3.
Weight vector 2	0.663	0.114	0.022	0.029	0.017	0.128	0.027	3.	2.
Weight vector 3	0.682	0.087	0.019	0.024	0.015	0.153	0.02	3.	4.

Table 3 The order of the selected tariffs and the weight vectors – Promethee II.

	Crit. 1	Crit. 2	Crit. 3	Crit. 4	Crit. 5	Crit. 6	Crit. 7
Type of generalized criteria	Level crit.	V-shape crit.	Usual crit.	Level crit.	Level crit.	Usual crit.	V-shape crit.
Parameter	d = 500 p = 2000	p = 500		d = 1 p = 3	d = 1.5 p = 2.5		p = 3

Table 4 Promethee II – preference function and parametres

5 Conclusion

We found out in this article, that it is possible to find such weight vectors that provide similar results as application of Monte Carlo simulation. We used two methods of multiple criteria decision making – TOPSIS and PROMETHEE II. a then we found such weight vectors, which can provide the same or similar results as Monte Carlo simulation.

It is clear, that there are very important weight vectors and choice of preference function and value of parametres.The position of alternative is influenced by this settings and these parametres are crucial. On the other hand If weights are unknown and cannot be determined, then Monte Carlo method might provide some information.

Acknowledgements

Supporting people and institutions (including Science Foundations) go here. The research project was supported by Grant No. IGA F4/16 /2011, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Evans, G.W.: An Overwiev of Techniques for Solving Multiobjective Mathematical Programs. *Management Science*. 1984, vol. 30, No. 11, 1268-1282.
[2] Fiala, P.: *Modely a metody rozhodování*. Praha: Oeconomica 2006.
[3] Figueira, J., Greco, S., Ehrgott M.: *Multiple Criteria Decision Analysis – State of the Art Surveys*. New York : Springer Science + Business Media Inc., 2005.
[4] Harák, T.: Využití simulací k redukci nákladů za telefonování v reálné firmě. Diploma Thesis, University of Economics, Prague 2012.
[5] Kuncová,M, Sekničková, J .: *Multi-criteria evaluation of alternatives applied to the mobile phone tariffs in comparison with monte carlo simulation results.* Bratislava: Quantitative Methods in Economics, 2012.
[6] Tomic V., Marinkovic Z., Janosevic D.: Promethee Method Implementation with Multiple-Criteria Decisions. *Mechanical Engineering.* 2011, vol. 9, No. 2, 193-202.

An algorithm for testing T5 solvability of max-plus interval systems

Helena Myšková[1]

Abstract. Max-plus algebra is the algebraic structure in which classical addition and multiplication are replaced by $a \oplus b = \max\{a, b\}$ and $a \otimes b = a + b$, respectively. Each system of linear equation in max-plus algebra we can write in the matrix form $A \otimes x = b$, where A and b are matrix and vector of suitable size. If we replace the matrix elements with matrix interval $\boldsymbol{A} = [\underline{A}, \overline{A}]$ and vector elements by vector interval $\boldsymbol{b} = [\underline{b}, \overline{b}]$, we get an interval system of linear equations. We can define several types of solvability of interval systems in max-plus algebra. In this paper, we shall deal with one of them, the so called T5 solvability. We give the algorithm which answers the question whether the given interval system is T5 solvable or not.

Keywords: max-plus algebra, interval system, T5 solvability

JEL classification: C02
AMS classification: 15A06; 65G30

1 Introduction

Problems on algebraic structures, in which pairs of operations $(\max, +)$ or $(\max, \min)$ replace addition and multiplication of the classical linear algebra, appear in the literature approximately since the sixties of the last century (see e.g. [3, 9]). A systematic theory of such algebraic structures was published probably for the first time in [3]. One of the problems we can deal with is solving of systems of linear equations, which are useful for modeling of discrete dynamic systems, scheduling or graph theory. Among interesting real-life applications let us mention, e.g., a large scale model of Dutch railway network or synchronizing traffic lights in Delft [7].

We describe in more detail one of possible applications in the following example, taken from [7], but slightly modified and generalized.

Example 1. There are two railway stations S_1 and S_2 in a metropolitan area, which are interconnected by a railway system consisting of two inner circles and two outer circles (see Figure 1). The number $a_{ij},\ i, j \in \{1, 2\}$ indicates the transit time from station S_j to station S_i including the time necessary for passengers to change over.

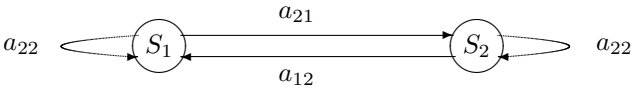


Figure 1

Suppose that there are four trains (two at each station) and two of them in the same station S_i leave simultaneously at the time x_i. The time at which both trains are already in station S_i is equal to $\max\{a_{i1} + x_1, a_{i2} + x_2\}$. Suppose that there are two schools near the two stations that begin their daily programme at the times b_1, b_2. It is required to find the departure times x_i which allow the students to catch the beginning of classes, i.e.,

$$\max_j (a_{ij} + x_j) = b_i \tag{1}$$

[1]Technical University in Košice/Faculty of Electrical Engineering and Informatics, Department of Mathematics, B. Nemcovej 32, 04001 Košice, Slovak Republic, e-mail: helena.myskova@tuke.sk

for $i = 1, 2$. Using the symbols $\oplus$ and $\otimes$ for operations of maximum and addition, respectively, we can rewrite (1) to the matrix form

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \otimes \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}. \tag{2}$$

The above described model system can be easily rewritten in general case of n stations. If there is no traffic from station S_j to station S_i we set $a_{ij} = -\infty$.

2 Preliminaries

By *max-plus algebra* we understand a triple $(B, \oplus, \otimes)$, where

$$B = \mathbb{R} \cup \{\varepsilon\},\ a \oplus b = \max\{a, b\},\ a \otimes b = a + b,$$

and $\varepsilon = -\infty$. Denote by M and N the index sets $\{1, 2, \ldots, m\}$ and $\{1, 2, \ldots, n\}$, respectively. The set of all $m \times n$ *matrices* over B is denoted by $B(m, n)$ and the set of all column n-*vectors* over B by $B(n)$. Operations $\oplus$ and $\otimes$ are extended to matrices and vectors in the same way as in classical algebra. We shall consider the *ordering* $\leq$ on the sets $B(m, n)$ and $B(n)$ defined as follows:

- for $A, C \in B(m, n):\ A \leq C$ if $a_{ij} \leq c_{ij}$ for all $i \in M,\ j \in N$,
- for $x, y \in B(n):\ x \leq y$ if $x_j \leq y_j$ for all $j \in N$.

We shall use the *monotonicity of* $\otimes$ which means that for each $A, C \in B(m, n)$ and for each $x, y \in B(n)$ the implication

$$\text{if } A \leq C \text{ and } x \leq y \text{ then } A \otimes x \leq C \otimes y$$

holds true.

By generalizing of (2) we get a *max-plus system of linear equations* of the form

$$A \otimes x = b \tag{3}$$

where $A \in B(m, n)$, $b \in B(m)$.

To give a necessary and sufficient condition for solvability of (3), we add some conditions (for details see [6]). We shall suppose that

i) $b_i > \varepsilon$ for all $i \in M$,
ii) A contains no column with full ε-s.

By now, we can define a principal solution of system (3) as follows:

$$x_j^*(A, b) = \min_{i \in M}\{b_i - a_{ij}\} \tag{4}$$

for each $j \in N$.

The following assertions describe the importance of the principal solution for the solvability of (3).

Lemma 1. [3, 10] Let $A \in B(m, n)$ and $b \in B(m)$ be given.

i) If $A \otimes x = b$ for $x \in B(n)$, then $x \leq x^*(A, b)$.
ii) $A \otimes x^*(A, b) \leq b$.

Theorem 1. [3] *Let $A \in B(m, n)$ and $b \in B(m)$ be given. Then the system $A \otimes x = b$ is solvable if and only if $x^*(A, b)$ is its solution.*

Lemma 2. [4] Let $A \in B(m, n),\ b, d \in B(m)$ be such that $b \leq d$. Then $x^*(A, b) \leq x^*(A, d)$.

Lemma 3. [4] Let $b \in B(m),\ C, D \in B(m, n)$ be such that $D \leq C$. Then $x^*(C, b) \leq x^*(D, b)$.

3 Interval systems

In practice, the traveling times in Example 1 may depend on outside conditions, so the values a_{ij} are from intervals of possible values, i.e., $a_{ij} \in [\underline{a}_{ij}, \overline{a}_{ij}]$ for each $i \in N,\ j \in N$. Also we shall require the arrival times to be not precise values but they are rather from given intervals, i.e., $b_i \in [\underline{b}_i, \overline{b}_i]$ for each $i \in N$.

Similarly to [1, 4, 5, 8] we define an *interval matrix* $\boldsymbol{A}$ and *interval vector* $\boldsymbol{b}$ as follows:

$$\boldsymbol{A} = [\underline{A}, \overline{A}] = \left\{ A \in B(m,n);\ \underline{A} \leq A \leq \overline{A} \right\}$$

and

$$\boldsymbol{b} = [\underline{b}, \overline{b}] = \left\{ b \in B(m);\ \underline{b} \leq b \leq \overline{b} \right\},$$

where $\underline{A}, \overline{A} \in B(m,n)$, $\underline{A} \leq \overline{A}$ and $\underline{b}, \overline{b} \in B(m)$, $\underline{b} \leq \overline{b}$.

Denote by

$$\boldsymbol{A} \otimes x = \boldsymbol{b} \tag{5}$$

the set of all max-plus systems of linear equations of the form (3) such that $A \in \boldsymbol{A},\ b \in \boldsymbol{b}$. We shall call (5) a *max-plus interval system of linear equations.*

A system of the form (3) is called a *subsystem* of (5) if $A \in \boldsymbol{A}$, $b \in \boldsymbol{b}$. We say, that interval system (5) has the *constant matrix* if $\underline{A} = \overline{A}$ and has the *constant right-hand side*, if $\underline{b} = \overline{b}$. Subsystem (3) is *extremal*, if each of the equations has the form $[\underline{A} \otimes x]_i = \overline{b}_i$ or $[\overline{A} \otimes x]_i = \underline{b}_i$ and we call them an LU *equation* or an *UL equation*, respectively.

To use the arguments from the previous section we shall suppose for interval system (5) that

i) $\underline{b}_i \neq \varepsilon$ for each $i \in M$,
ii) for each $j \in N$ there exists $i \in M$ such that $\underline{a}_{ij} \neq \varepsilon$.

We can define several conditions which the given interval system is required to fulfill. According to them we shall define several solvability concepts. Table 1 contains the list of all up to now studied types of the solvability of (5) in max-plus algebra. There are omitted solvability concepts which lead to trivial conditions.

Solvability concept	**Definition**
Weak solvability [1]	$(\exists x \in B(n))(\exists A \in \boldsymbol{A})(\exists b \in \boldsymbol{b})\colon A \otimes x = b$
Strong solvability [2]	$(\forall A \in \boldsymbol{A})(\forall b \in \boldsymbol{b})(\exists x \in B(n))\colon A \otimes x = b$
Tolerance solvability [1]	$(\exists x \in B(n))(\forall A \in \boldsymbol{A})(\exists b \in \boldsymbol{b})\colon A \otimes x = b$
Weak tolerance solvability [4]	$(\forall A \in \boldsymbol{A})(\exists x \in B(n))(\exists b \in \boldsymbol{b})\colon A \otimes x = b$
Control solvability [5]	$(\exists x \in B(n))(\forall b \in \boldsymbol{b})(\exists A \in \boldsymbol{A})\colon A \otimes x = b$
Weak control solvability [5]	$(\forall b \in \boldsymbol{b})(\exists x \in B(n))(\exists A \in \boldsymbol{A})\colon A \otimes x = b$
Universal solvability [4]	$(\exists x \in B(n))(\forall b \in \boldsymbol{b})(\forall A \in \boldsymbol{A})\colon A \otimes x = b$
Weak universal solvability [5]	$(\forall b \in \boldsymbol{b})(\exists x \in B(n))(\forall A \in \boldsymbol{A})\colon A \otimes x = b$
T4 solvability [6]	$(\exists b \in \boldsymbol{b})(\exists x \in B(n))(\forall A \in \boldsymbol{A})\colon A \otimes x = b$

Table 1: Solvability concepts

4 T5 solvability

The notions of a T5-vector and the T5 solvability of interval system (5) are defined in this section. The procedure for checking the T5 solvability is presented.

Definition 1.

i) A vector $b \in \boldsymbol{b}$ is called a T5-vector of interval system (5), if for each $A \in \boldsymbol{A}$ there exists $x \in B(n)$ such that $A \otimes x = b$.
ii) Interval system (5) is T5 solvable if there exists a vector $b \in \boldsymbol{b}$ such that b is its T5-vector.

To give a necessary and sufficient condition for the T5 solvability we recall the notion of the *strong solvability*, which has been studied in max-plus and max-min algebra by K. Cechlárová and R. Cuninghame-Green [2].

Definition 2. Interval system (5) is strongly solvable if for each $b \in \boldsymbol{b}$ and for each $A \in \boldsymbol{A}$ there exists $x \in B(n)$ such that $A \otimes x = b$.

Theorem 2. [2] *Interval system (5) is strongly solvable if and only if all its extremal subsystems with exactly one LU equation are solvable.*

For each $k = 1, 2, \ldots, m$ denote by $A^{(k)}$ the matrix with entries

$$a_{ij}^{(k)} = \begin{cases} \underline{a}_{ij} & \text{for} \quad i = k,\ j \in N, \\ \overline{a}_{ij} & \text{for} \quad i \neq k,\ j \in N. \end{cases}$$

Lemma 4. A vector $b \in \boldsymbol{b}$ is a T5-vector of interval system (5) if and only if

$$A^{(k)} \otimes x^*(A^{(k)}, b) = b \tag{6}$$

for each $k \in M$.

Proof. A vector $b \in \boldsymbol{b}$ is a T5-vector of interval system (5) if and only if interval system (5) with the constant right-hand side $\underline{b} = \overline{b} = b$ is strongly solvable. By Theorem 1 and Theorem 2, the strong solvability of an interval system (5) with $\underline{b} = \overline{b} = b$ is equivalent to the validity of (6) for each $k \in M$. □

This lemma does not give a method for finding a T5-vector. For this reason, we define a *T5-sequence* of interval system (5).

Definition 3. The T5-sequence of interval system (5) is a sequence $\{c^{(k)}\}_{k=0}^{\infty}$ defined as follows:

$$\begin{aligned} c^{(0)} &= \overline{b}, \\ c_i^{(k+1)} &= \min_{r \in M} \{[A^{(r)} \otimes x^*(A^{(r)}, c^{(k)})]_i\}, \end{aligned} \tag{7}$$

for each $k \in \mathbb{N}_0$, $i \in M$.

Theorem 3. *Let $c \in \boldsymbol{b}$ be a T5-vector of interval system (5). Then for each nonnegative integer k the inequality $c \leq c^{(k)}$ is satisfied.*

Proof. By mathematical induction on k

1. For $k = 0$ the inequality $c \leq \overline{b} = c^{(0)}$ is trivially satisfied.
2. We prove that if $c \leq c^{(k)}$ then $c \leq c^{(k+1)}$.
   For the sake of contradiction suppose that $c \leq c^{(k)}$ and there exists $i \in M$ such that $c_i > c_i^{(k+1)}$. Using Lemma 2 we get

$$c_i > \min_{r \in M} \{[A^{(r)} \otimes x^*(A^{(r)}, c^{(k)})]_i\} \geq \min_{r \in M} \{[A^{(r)} \otimes x^*(A^{(r)}, c)]_i\},$$

   which implies that there exists $p \in M$ such that $A^{(p)} \otimes x^*(A^{(p)}, c) \neq c$. By Lemma 4 the vector c is not a T5-vector of (5), a contradiction. □

Lemma 5. Let $\{c^{(k)}\}_{k=0}^{\infty}$ be the T5-sequence of interval system(5) and $l \in \mathbb{N}_0$ be arbitrary. The following assertions hold true:

i) The sequence $\{c^{(k)}\}_{k=0}^{\infty}$ is non-increasing.
ii)A vector $c^{(l)} \in \boldsymbol{b}$ is a T5-vector of (5) if and only if $c^{(l+1)} = c^{(l)}$.

Proof.

i) By Lemma 1b) we have $[A^{(r)} \otimes x^*(A^{(r)}, c^{(k)})]_i \leq c_i^{(k)}$ for each $r \in M$, $i \in M$ which implies $c_i^{(k+1)} = \min\limits_{r \in M}\{[A^{(r)} \otimes x^*(A^{(r)}, c^{(k)})]_i\} \leq c_i^{(k)}$ for each $i \in M$, so the sequence $\{c^{(k)}\}_{k=0}^{\infty}$ is non-increasing.

ii) According to Lemma , a vector $c^{(l)}$ is a T5-vector of (5) if and only if $[A^{(r)} \otimes x^*(A^{(r)}, c^{(l)})]_i = c_i^{(l)}$ for each $r \in M$, $i \in M$ which is equivalent to $c_i^{(l+1)} = \min\limits_{r \in M}\{[A^{(r)} \otimes x^*(A^{(r)}, c^{(l)})]_i\} = c_i^{(l)}$. □

Now, we can suggest the algorithm for checking the T5 solvability.

Algorithm T5

Input: $\boldsymbol{A}$,$\boldsymbol{b}$

Output: 'yes' in variable $t5$ if the given interval system is T5 solvable, and 'no' in $t5$ otherwise

Step 1. $c^{(0)} = \overline{b},\ k = 0$

Step 2. For each $i \in M$ compute $c_i^{(k+1)} = \min\limits_{r \in M}\{[A^{(r)} \otimes x^*(A^{(r)}, c^{(k)})]_i\}$

Step 3. If $\underline{b} \not\leq c^{(k+1)}$ then $t5 := no$, go to **end**
Step 4. If $c^{(k+1)} = c^{(k)}$ then $t5 := yes$, $c^* = c^{(k)}$ go to **end**
Step 5. $k = k + 1$, go to *Step 2*
end

Remark 1. If interval system (5) in the max-plus algebra is T5 solvable then vector c^* is its maximal T5-vector.

Remark 2. Using Algorithm T5 for the model system described in Example 1 we can find the vector of arrival times $c^* \in \boldsymbol{b}$ which can be achieved by suitable choice of the vector x of the departure times depending of the transit times between the stations, if such a vector of arrival times exists.

Example 2. Check the T5 solvability of the interval system $\boldsymbol{A} \otimes x = \boldsymbol{b}$, where

$$\boldsymbol{A} = \begin{pmatrix} [16,17] & [5,12] & [8,10] \\ [7,13] & [11,12] & [5,5] \\ [4,8] & [1,3] & [15,15] \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} [20,22] \\ [15,18] \\ [14,17] \end{pmatrix}.$$

We have

$$A^{(1)} = \begin{pmatrix} 16 & 5 & 8 \\ 13 & 12 & 5 \\ 8 & 3 & 15 \end{pmatrix},\ A^{(2)} = \begin{pmatrix} 17 & 12 & 10 \\ 7 & 11 & 5 \\ 8 & 3 & 15 \end{pmatrix},\ A^{(3)} = \begin{pmatrix} 17 & 12 & 10 \\ 13 & 12 & 5 \\ 4 & 1 & 15 \end{pmatrix}.$$

and $c^{(0)} = (22, 18, 17)^T$. We compute

$$x^*(A^{(1)}, c^{(0)}) = \begin{pmatrix} 5 \\ 6 \\ 2 \end{pmatrix}, \quad A^{(1)} \otimes x^*(A^{(1)}, c^{(0)}) = \begin{pmatrix} 21 \\ 18 \\ 17 \end{pmatrix},$$

$$x^*(A^{(2)}, c^{(0)}) = \begin{pmatrix} 5 \\ 7 \\ 2 \end{pmatrix}, \quad A^{(2)} \otimes x^*(A^{(2)}, c^{(0)}) = \begin{pmatrix} 22 \\ 18 \\ 17 \end{pmatrix},$$

$$x^*(A^{(3)}, c^{(0)}) = \begin{pmatrix} 5 \\ 6 \\ 2 \end{pmatrix}, \quad A^{(3)} \otimes x^*(A^{(3)}, c^{(0)}) = \begin{pmatrix} 22 \\ 18 \\ 17 \end{pmatrix}.$$

By (7), we get $c^{(1)} = (21, 18, 17)^T$ and consequently we compute

$$x^*(A^{(1)}, c^{(1)}) = \begin{pmatrix} 5 \\ 6 \\ 2 \end{pmatrix}, \quad A^{(1)} \otimes x^*(A^{(1)}, c^{(1)}) = \begin{pmatrix} 21 \\ 18 \\ 17 \end{pmatrix} = c^{(1)},$$

$$x^*(A^{(2)}, c^{(1)}) = \begin{pmatrix} 4 \\ 7 \\ 2 \end{pmatrix}, \quad A^{(2)} \otimes x^*(A^{(2)}, c^{(1)}) = \begin{pmatrix} 21 \\ 18 \\ 17 \end{pmatrix} = c^{(1)},$$

$$x^*(A^{(3)}, c^{(1)}) = \begin{pmatrix} 4 \\ 6 \\ 2 \end{pmatrix}, \quad A^{(3)} \otimes x^*(A^{(3)}, c^{(1)}) = \begin{pmatrix} 21 \\ 18 \\ 17 \end{pmatrix} = c^{(1)}.$$

Since $A^{(k)} \otimes x^*(A^{(k)}, c^{(1)}) = c^{(1)}$ for $k = 1, 2, 3$, the given interval system is T5 solvable with the vector $c^{(1)} = c^*$ as the maximal T5-vector.

References

[1] Cechlárová, K.: Solutions of interval systems in max-plus algebra. In: *Proc. SOR 2001* (V. Rupnik, L. Zadnik-Stirn, S. Drobne, eds.), Preddvor, Slovenia, 321–326.

[2] Cechlárová, K., and Cuninghame-Green, R. A.: Interval systems of max-separable linear equations, *Linear Algebra and its Applications* **340** (2002), 215–224.

[3] Cuninghame-Green, R. A.: Minimax Algebra. *Lecture notes in Economics and Mathematical systems* 1966, Springer, Berlin, 1979.

[4] Myšková, H.: Interval systems of max-separable linear equations, *Linear Algebra and its Applications* **403** (2005), 263–272.

[5] Myšková, H.: Control solvability of interval systems of max-separable linear equations, *Linear Algebra and its Applications* **416** (2006), 215–223.

[6] Myšková, H.: On an algorithm for testing T4 solvability of max-plus interval systems, *Kybernetika*, article in press.

[7] G. J. Oldser et al.: Course Notes: Max-algebra Aproach to Discrete Event Systems, *Algebres Max-Plus et Applications an Informatique et Automatique*, INRIA, 1998, 147-196.

[8] Gavalec, M., and Plavka, J.: Monotone interval eigenproblem in max-min algebra, *Kybernetika* **46** (3) (2010), 387-396.

[9] Vorobjov, N. N.: Extremal Algebra of positive Matrices, *Datenverarbeitung und Kybernetik* **3** (1967), 39-71 (in Russian).

[10] Zimmermann, K.: Extremální algebra. *Ekonomicko-matematická laboratoř Ekonomického ústavu ČSAV*, Praha 1976.

A simulation study on an approximate confidence region of parameters of a quadratic calibration function

Kateřina Myšková[1]

Abstract. The goal of the paper is to derive the approximate confidence region of parameters of a quadratic calibration function in a one-dimensional replicated calibration model and to verify the accuracy of this approximate confidence region by a simulation study. We view calibration as a method for describing a relationship between two imprecise measurements. The relationship is called calibration function, we suppose this function in a quadratic form. There are two approximations in deriving the confidence region. First is replacement the nonlinear model by a linear one using a first-order Taylor series. Second is using Kenward-Roger approach for approximation the variance-covariance matrix of the estimators of parameters of the calibration function. The simulation study will be performed in the computer system Matlab. It will be concentrated on the validity of expression of the approximate confidence region depending on the number of measurement objects, on the number of replications and on the covariance matrix.

Keywords: one-dimensional calibration model, quadratic calibration function, MINQUE method, Kenward-Roger approach, simulation study.

JEL classification: C15
AMS classification: 62J05

1 Introduction

We view calibration as a relationship between two measurements provided that both are imprecise. Sometimes, such a calibration is called comparative. The main purpose of calibration is to estimate the unknown parameters of a function describing such a relationship, called a calibration function.

Assuming a calibration function to be quadratic is a natural generalization of the assumption of a linear one motivated also by practical reasons since there may be situations in which a quadratic function describes the relationship better.

Moreover replicating a measurement makes it possible to estimate the unknown parameters of a model's covariance matrix. To this end, the paper uses a MINQUE method. An estimate of the covariance matrix is not sufficient for establishing a confidence region for the parameters to be estimated. However, Kenward-Roger's procedure can be used to derive a Wald-type statistic along with its approximate F-distribution to construct an approximate confidence region.

This paper contains two parts. First part is concetred on deriving the approximate confidence region. Second part describes results of a small simulation study, which should verify the validity of the aproximate confidence region.

Our motivation for this model was to calibrate two measurements of moisture in corn. We have two replicated measurements from two methods, one is standard and one is quicker. We consider a quadratic calibration function because we know that both devices measure well on a scale from 10 % to 15 % and anything below and above is less precise. We can verify this relation by testing the significance of the parameters of the calibration function.

[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 Deriving a model

Suppose that a single value is measured for n objects using two different methods each time with a series of m independent measurement made. We may also assume independence of the first and second measurement series and of the measurements of individual objects.

Let then the result of the j-th repetition of the first-series measurement is the realization of an n-dimensional random vector for which

$$\mathbf{X}^j \sim N_n(\boldsymbol{\mu}, \sigma_x^2 \mathbf{I}_n) \text{ for } j = 1, \ldots, m.$$

The result of the j-th repetition of the second-series measurement is the realization of an n-dimensional random vector for which

$$\mathbf{Y}^j \sim N_n(\boldsymbol{\nu}, \sigma_y^2 \mathbf{I}_n) \text{ for } j = 1, \ldots, m.$$

The results of all the measurement may be summarized as

$$\begin{pmatrix} \mathbf{X}^1 \\ \mathbf{Y}^1 \\ \vdots \\ \mathbf{X}^m \\ \mathbf{Y}^m \end{pmatrix} \sim N\left[\mathbf{1}_m \otimes \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\nu} \end{pmatrix}, \mathbf{I}_m \otimes \begin{pmatrix} \sigma_x^2 \mathbf{I}_n & \mathbf{0}_n \\ \mathbf{0}_n & \sigma_y^2 \mathbf{I}_n \end{pmatrix} \right].$$

Further, let us assume that the relationship between the actual (error-free) measurement values (the calibration function) is quadratic:

$$\boldsymbol{\nu} = a\mathbf{1}_n + b\boldsymbol{\mu} + c\,\mathrm{Diag}^2(\boldsymbol{\mu})\mathbf{1}_n,$$

where $a, b, c \in R$, $\mathrm{Diag}(\boldsymbol{\mu})$ is a diagonal matrix with entries on the diagonal given by the components of vector $\boldsymbol{\mu}$, $\mathbf{1}_n = (1, \ldots, 1)' \in R^n$.

As the calibration function is non-linear (involving products of the unknown parameters), we linearize it using a Taylor series expansion around b_0, c_0, $\boldsymbol{\mu}_0$. Neglecting terms of order two and higher, we have

$$\boldsymbol{\nu} \doteq \mathbf{1}_n a + \boldsymbol{\mu}_0 b + \mathrm{Diag}^2(\boldsymbol{\mu_0})\mathbf{1}_n c + \mathrm{Diag}(b_0\mathbf{1}_n + 2c_0\boldsymbol{\mu}_0)\delta\boldsymbol{\mu},$$

where $\delta\boldsymbol{\mu} = \boldsymbol{\mu} - \boldsymbol{\mu}_0$. In terms of the vectors of unknown parameters, the calibration function may be rewritten as

$$(\mathrm{Diag}(b_0\mathbf{1}_n + 2c_0\boldsymbol{\mu}_0), -\mathbf{I}_n) \begin{pmatrix} \delta\boldsymbol{\mu} \\ \boldsymbol{\nu} \end{pmatrix} + (\mathbf{1}_n, \boldsymbol{\mu}_0, \mathrm{Diag}^2(\boldsymbol{\mu_0})\mathbf{1}_n) \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \mathbf{0}$$

and this form thought of as conditions for the unknown parameters.

Thus, we get the model

$$\begin{pmatrix} \mathbf{X}^1 - \boldsymbol{\mu}_0 \\ \mathbf{Y}^1 \\ \vdots \\ \mathbf{X}^m - \boldsymbol{\mu}_0 \\ \mathbf{Y}^m \end{pmatrix} \sim N\left[\mathbf{1}_m \otimes \begin{pmatrix} \delta\boldsymbol{\mu} \\ \boldsymbol{\nu} \end{pmatrix}, \mathbf{I}_m \otimes \begin{pmatrix} \sigma_x^2 \mathbf{I}_n & \mathbf{0}_n \\ \mathbf{0}_n & \sigma_y^2 \mathbf{I}_n \end{pmatrix} \right]$$

with the following parameter conditions

$$(\mathrm{Diag}(b_0\mathbf{1}_n + 2c_0\boldsymbol{\mu}_0), -\mathbf{I}_n) \begin{pmatrix} \delta\boldsymbol{\mu} \\ \boldsymbol{\nu} \end{pmatrix} + (\mathbf{1}_n, \boldsymbol{\mu}_0, \mathrm{Diag}^2(\boldsymbol{\mu_0})\mathbf{1}_n) \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \mathbf{0}.$$

We call this model a one-dimensional replicated calibration model with a quadratic calibration function.

2.1 Estimators of model parameters

Next we will be concerned with expressing estimators of the unknown parameters using this model. Note that estimators of the unknown parameters in a one-dimensional non-replicated calibration model ($m = 1$) with a quadratic calibration function can be found in [3].

Supposing known covariance matrix (known parameters σ_x^2, σ_y^2), denoting $\mathbf{D} = \mathrm{Diag}(b_0\mathbf{1}_n + 2c_0\boldsymbol{\mu}_0)$, $\mathbf{A} = (\mathbf{1}_n, \boldsymbol{\mu}_0, \mathrm{Diag}^2(\boldsymbol{\mu_0})\mathbf{1}_n)$, $\mathbf{W} = \sigma_x^2\mathbf{D}^2 + \sigma_y^2\mathbf{I}_n$ and using the relationships found in [1, p. 129], we will write the estimators:

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}} - \sigma_x^2\mathbf{D}(\mathcal{P}_\mathbf{A}\mathbf{W}\mathcal{P}_\mathbf{A})^+(\mathbf{D}(\bar{\mathbf{X}} - \boldsymbol{\mu}_0) - \bar{\mathbf{Y}}),$$

$$\hat{\boldsymbol{\nu}} = \bar{\mathbf{Y}} + \sigma_y^2(\mathcal{P}_\mathbf{A}\mathbf{W}\mathcal{P}_\mathbf{A})^+(\mathbf{D}(\bar{\mathbf{X}} - \boldsymbol{\mu}_0) - \bar{\mathbf{Y}}),$$

$$\begin{pmatrix}\hat{a}\\ \hat{b}\\ \hat{c}\end{pmatrix} = -(\mathbf{A}'\mathbf{W}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{W}^{-1}(\mathbf{D}(\bar{\mathbf{X}} - \boldsymbol{\mu}_0) - \bar{\mathbf{Y}}),$$

where $\mathcal{P}_\mathbf{A}$ is projection matrix on the orthogonal complement of the columns of matrix $\mathbf{A}$ and $(\mathcal{P}_\mathbf{A}\mathbf{W}\mathcal{P}_\mathbf{A})^+ = \mathbf{W}^{-1} - \mathbf{W}^{-1}\mathbf{A}(\mathbf{A}'\mathbf{W}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{W}^{-1}$ (a Moore-Penrose pseudoinverse). The covariance matrices of the estimators are

$$\mathrm{var}\begin{pmatrix}\hat{\boldsymbol{\mu}}\\ \hat{\boldsymbol{\nu}}\end{pmatrix} = \frac{1}{m}\begin{pmatrix}\sigma_x^2\mathbf{I}_n - \sigma_x^4\mathbf{D}(\mathcal{P}_\mathbf{A}\mathbf{W}\mathcal{P}_\mathbf{A})^+\mathbf{D} & \sigma_x^2\sigma_y^2\mathbf{D}(\mathcal{P}_\mathbf{A}\mathbf{W}\mathcal{P}_\mathbf{A})^+\\ \sigma_x^2\sigma_y^2(\mathcal{P}_\mathbf{A}\mathbf{W}\mathcal{P}_\mathbf{A})^+\mathbf{D} & \sigma_y^2\mathbf{I}_n - \sigma_y^4(\mathcal{P}_\mathbf{A}\mathbf{W}\mathcal{P}_\mathbf{A})^+\end{pmatrix},$$

$$\mathrm{var}\begin{pmatrix}\hat{a}\\ \hat{b}\\ \hat{c}\end{pmatrix} = \frac{1}{m}(\mathbf{A}'\mathbf{W}^{-1}\mathbf{A})^{-1},$$

$$\mathrm{cov}\left(\begin{pmatrix}\hat{a}\\ \hat{b}\\ \hat{c}\end{pmatrix}, \begin{pmatrix}\hat{\boldsymbol{\mu}}\\ \hat{\boldsymbol{\nu}}\end{pmatrix}\right) = -\frac{1}{m}(\mathbf{A}'\mathbf{W}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{W}^{-1}(\sigma_x^2\mathbf{D}, -\sigma_y^2\mathbf{I}_n).$$

Note that if the covariance matrix is unknown and we have estimates of the parameters of this matrix, the estimators of the model are in the same form. However, the estimates $\hat{\sigma}_x^2$, $\hat{\sigma}_y^2$ are used instead of the true values σ_x^2, σ_y^2. One possibility to receive the estimates is the MINQUE method described in the part 3. In the last part of this paper, we will construct a approximate confidence region for the vector of the parameters of a calibration function, if the parameters of the covariance matrix are unkown.

2.2 Estimator of covariance matrix parameters

As mentioned above, using a replicated model, also estimates may be expressed of the covariance matrix unknown parameters. This can be done using a MINQUE method, see, for example, [1] or [5] for more information. This method has a local character requiring initial estimates. The covariance matrix of our model depends on two parameters only, σ_x^2, σ_y^2. Denote $\sigma_{x,0}^2$, $\sigma_{y,0}^2$ the initial estimates of the unknown parameters and $\mathbf{W}_0 = \sigma_{x,0}^2\mathbf{D}^2 + \sigma_{y,0}^2\mathbf{I}_n$. Then the $(\sigma_{x,0}^2, \sigma_{y,0}^2)'$-MINQUE estimate of the vector $(\sigma_x^2, \sigma_y^2)'$ is given by the equation

$$\begin{pmatrix}\hat{\sigma}_x^2\\ \hat{\sigma}_y^2\end{pmatrix} = \left[n(m-1)\begin{pmatrix}\sigma_{x,0}^{-4} & 0\\ 0 & \sigma_{y,0}^{-4}\end{pmatrix} + \mathbf{S}\right]^{-1}\begin{pmatrix}\gamma_1\\ \gamma_2\end{pmatrix},$$

where

$$\gamma_1 = \sigma_{x,0}^{-4}\sum_{k=1}^{m}(\mathbf{X}^k - \overline{\mathbf{X}})'(\mathbf{X}^k - \overline{\mathbf{X}}) + m(\overline{\mathbf{X}} - \hat{\boldsymbol{\mu}})'(\overline{\mathbf{X}} - \hat{\boldsymbol{\mu}}),$$

$$\gamma_2 = \sigma_{y,0}^{-4}\sum_{k=1}^{m}(\mathbf{Y}^k - \overline{\mathbf{Y}})'(\mathbf{Y}^k - \overline{\mathbf{Y}}) + m(\overline{\mathbf{Y}} - \hat{\boldsymbol{\nu}})'(\overline{\mathbf{Y}} - \hat{\boldsymbol{\nu}}),$$

$$\overline{\mathbf{X}} = \frac{1}{m}\sum_{k=1}^{m}\mathbf{X}^k, \quad \overline{\mathbf{Y}} = \frac{1}{m}\sum_{k=1}^{m}\mathbf{Y}^k$$

and

$$\mathbf{S} = \begin{pmatrix} \mathrm{Tr}\left\{(\mathbf{D}(\mathcal{P}_{\mathbf{A}}\mathbf{W}_0\mathcal{P}_{\mathbf{A}})^{+}\mathbf{D})^2\right\} & \mathrm{Tr}\left\{(\mathbf{D}(\mathcal{P}_{\mathbf{A}}\mathbf{W}_0\mathcal{P}_{\mathbf{A}})^{+})^2\right\} \\ \mathrm{Tr}\left\{((\mathcal{P}_{\mathbf{A}}\mathbf{W}_0\mathcal{P}_{\mathbf{A}})^{+}\mathbf{D})^2\right\} & \mathrm{Tr}\left\{((\mathcal{P}_{\mathbf{A}}\mathbf{W}_0\mathcal{P}_{\mathbf{A}})^{+})^2\right\} \end{pmatrix}.$$

The estimator's covariance matrix is

$$\mathrm{var}\begin{pmatrix}\hat{\sigma}_x^2 \\ \hat{\sigma}_y^2\end{pmatrix} = 2\left[n(m-1)\begin{pmatrix}\sigma_{x,0}^{-4} & 0 \\ 0 & \sigma_{y,0}^{-4}\end{pmatrix} + \mathbf{S}\right]^{-1}.$$

2.3 Approximate confidence region

We will use a procedure by Kenward and Roger (see [2]) to construct an approximate confidence region for the calibration function's parameter vector applying it to the model

$$\begin{pmatrix}\hat{a} \\ \hat{b} \\ \hat{c}\end{pmatrix} \sim N\left[\begin{pmatrix}a \\ b \\ c\end{pmatrix}, \left(\mathbf{A}'\left(\frac{1}{m}\mathbf{W}\right)^{-1}\mathbf{A}\right)^{-1}\right].$$

Using the notation from [2] we denote the covariance matrix

$$\mathbf{\Phi} = \frac{1}{m}(\mathbf{A}'\mathbf{W}^{-1}\mathbf{A})^{-1}.$$

Based on the above paper, we will write an adjusted covariance matrix:

$$\hat{\mathbf{\Phi}}_A = \hat{\mathbf{\Phi}} + 2\hat{\mathbf{\Phi}}\{\sum_{x,y}\sum_{x,y}\mathbf{V}_{ij}(\mathbf{Q}_{ij} - \mathbf{P}_i\hat{\mathbf{\Phi}}\mathbf{P}_j)\}\hat{\mathbf{\Phi}},$$

where

$$\mathbf{V} = \begin{pmatrix}V_{xx} & V_{xy} \\ V_{yx} & V_{yy}\end{pmatrix} = 2\left[n(m-1)\begin{pmatrix}\sigma_{x,0}^{-4} & 0 \\ 0 & \sigma_{y,0}^{-4}\end{pmatrix} + \mathbf{S}\right]^{-1},$$

$$\mathbf{P}_x = -\frac{1}{m}\mathbf{A}'\mathbf{W}^{-1}\mathbf{D}^2\mathbf{W}^{-1}\mathbf{A}, \quad \mathbf{P}_y = -\frac{1}{m}\mathbf{A}'\mathbf{W}^{-1}\mathbf{W}^{-1}\mathbf{A},$$

$$\mathbf{Q}_{xx} = \frac{1}{m^2}\mathbf{A}'\mathbf{W}^{-1}\mathbf{D}^2\mathbf{W}^{-1}\mathbf{D}^2\mathbf{W}^{-1}\mathbf{A}, \quad \mathbf{Q}_{xy} = \frac{1}{m^2}\mathbf{A}'\mathbf{W}^{-1}\mathbf{D}^2\mathbf{W}^{-1}\mathbf{W}^{-1}\mathbf{A},$$

$$\mathbf{Q}_{yx} = \frac{1}{m^2}\mathbf{A}'\mathbf{W}^{-1}\mathbf{W}^{-1}\mathbf{D}^2\mathbf{W}^{-1}\mathbf{A}, \quad \mathbf{Q}_{yy} = \frac{1}{m^2}\mathbf{A}'\mathbf{W}^{-1}\mathbf{W}^{-1}\mathbf{W}^{-1}\mathbf{A}$$

and $\hat{\mathbf{\Phi}}$ has the same form as $\mathbf{\Phi}$, with the estimates $\hat{\sigma}_x^2$, $\hat{\sigma}_y^2$ calculated by the MINQUE method used for the unknown parameters σ_x^2, σ_y^2.

After some rather lengthy calculations we get

$$A_1 = \sum_{x,y}\sum_{x,y} V_{ij}\,\mathrm{Tr}\{\mathbf{P}_i\hat{\mathbf{\Phi}}\}\mathrm{Tr}\{\mathbf{P}_j\hat{\mathbf{\Phi}}\}, \quad A_2 = \sum_{x,y}\sum_{x,y} V_{ij}\,\mathrm{Tr}\{\mathbf{P}_i\hat{\mathbf{\Phi}}\mathbf{P}_j\hat{\mathbf{\Phi}}\},$$

$$B = \frac{1}{6}(A_1 + 6A_2),$$

$$g = \frac{4A_1 - 7A_2}{5A_2},$$

$$c_1 = \frac{4A_1 - 7A_2}{8A_1 + 31A_2}, \quad c_2 = \frac{-4A_1 + 22A_2}{8A_1 + 31A_2}, \quad c_3 = \frac{-4A_1 + 32A_2}{8A_1 + 31A_2},$$

$$E^* = (1 - \frac{A_2}{3})^{-1},$$

$$V^* = \frac{2}{3}\frac{1 + c_1B}{(1 - c_2B)^2(1 - c_3B)},$$

$$\rho = \frac{V^*}{2E^{*2}},$$
$$u = 4 + \frac{5}{3\rho - 1},$$
$$\lambda = \frac{u}{E^*(u-2)}.$$

Then we can derive that the λ-multiple of the statistic

$$F = \frac{1}{3}\left(\begin{pmatrix} \hat{a}-a \\ \hat{b}-b \\ \hat{c}-c \end{pmatrix}' \hat{\boldsymbol{\Phi}}_A^{-1} \begin{pmatrix} \hat{a}-a \\ \hat{b}-b \\ \hat{c}-c \end{pmatrix} \right)$$

has an approximate F-distribution with degrees of freedom of 3 and u. The approximate confidence region for the calibration function parameters vector is a set of vectors $(a, b, c)'$ which fulfill the inequality:

$$P(F \leq \frac{1}{\lambda} F_{3,u}(1-\alpha)) = 1 - \alpha. \tag{1}$$

3 Small simulation study

We verified the validity of the aproximate confidence interval (1) for the calibration function parameters vector $(a, b, c)'$ by simulations in software Matlab, in which we implemented the derived formulas. Values of the first-series and the second-series measurement have been generated at fixed values of the model parameters, particularly at fixed values of the first measurement mean values $\boldsymbol{\mu}$, quadratic calibration function parameters a, b , c, a number of replications m and variances σ_x^2, σ_y^2. We chose parameters of the calibration function as follows $a = 1$, $b = 2$, $c = 3$. For the number of replications m we selected numbers 3, 5 and 10. Standard deviations (variances) have been chosen in different ways (see below), as well as the mean values vector of the first measurement $\boldsymbol{\mu}$ has been chosen in several ways, but always elements of the vector were symmetrically and equidistantly around the parabola's minimum. As an estimate of the confidence, we calculated the ratio the generated values satisfying the inequality (1) to all generated values.

Standard deviations effect

For the standard deviations (variances) effect research, we considered the mean values vector of the first measurement $\boldsymbol{\mu} = (-10, -9, .., 8, 9)'$ and the number of replications $m = 3$. For each choice of two standard deviations we carried out 10 000 repetitions.

$\sigma_x = \sigma_y$	$\sigma_x = 0.05$ $\sigma_y = 0.05$ 94.12 %	$\sigma_x = 0.01$ $\sigma_y = 0.01$ 94.01 %	$\sigma_x = 0.1$ $\sigma_y = 0.1$ 93.59 %	$\sigma_x = 0.5$ $\sigma_y = 0.5$ 94.55 %			94.07 %
$\sigma_x < \sigma_y$	$\sigma_x = 0.01$ $\sigma_y = 0.05$ 94.32 %	$\sigma_x = 0.01$ $\sigma_y = 0.1$ 94.72 %	$\sigma_x = 0.01$ $\sigma_y = 0.5$ 96.66 %	$\sigma_x = 0.1$ $\sigma_y = 0.5$ 94.06 %	$\sigma_x = 0.05$ $\sigma_y = 0.1$ 94.56 %	$\sigma_x = 0.05$ $\sigma_y = 0.5$ 94.41 %	94.79 %
$\sigma_x > \sigma_y$	$\sigma_x = 0.1$ $\sigma_y = 0.05$ 93.38 %	$\sigma_x = 0.5$ $\sigma_y = 0.05$ 87.83 %	$\sigma_x = 0.5$ $\sigma_y = 0.1$ 88.64 %				89.95 %

Table 1 The simulation results of standard deviations effect

The best results are when the first measurement standard deviation is less than the standard deviation of the second measurement ($\sigma_x < \sigma_y$) and reaches almost 95 %. We can say, that the bigger differences the better results are. Percentage of the confidence for the same values of standard deviations is around 94 %. If $\sigma_x > \sigma_y$, the confidence is only about 90 %. The obtained results may cause the fact that the values of the second measurement are from a wider interval than the values of the first measurement.

Effect of replications

Effect of measurement replications, we considered for the the mean values vector of the first measurement $\boldsymbol{\mu} = (-10, -9, .., 8, 9)'$ and for four cases of standard deviations $\sigma_x = 0.1$, $\sigma_y = 0.5$; $\sigma_x = 0.5$, $\sigma_y = 0.5$; $\sigma_x = 0.01$, $\sigma_y = 0.5$; $\sigma_x = 0.1$, $\sigma_y = 0.1$. For all cases, we carried out 10 000 repetitions and we got results:

$m = 3$	$m = 5$	$m = 10$
94.43 %	94.61 %	94.69 %

Table 2 The simulation results of replications effect

As you can see in the table 2, the confidence is around 94.5 %, probably due to we chose only the cases of two variances, for which we achieved "good" results in the previous part. With a growing number of replications increases the confidence, but not particularly striking.

Number of measument objects effect

In the last part we focuse on the effect of measurement objects number. We suppose that all objects (points) are located equidistantly and symmetrically around the minimum of the parabola. We set values of other parameters as $\sigma_x = 0.1$, $\sigma_y = 0.5$ and $\sigma_x = 0.5$, $\sigma_y = 0.5$ for standard deviations and $m = 10$ for the number of replications.

$n = 10$	$n = 20$	$n = 40$
94.05 %	94.39 %	94,48 %

Table 3 The simulation results of effect of measurement objects number

The resulting values show that the increasing number of measurement objects cause the increasing confidence. All the same the confidence only get closer to 95 %. The diference between confidences for 10 and 20 measurement objects are stronger than the diference of confidences between 20 and 40 measurement objects.

4 Conclusion

Globally we could say that the confidence obtained by the simulation is approximatly around 94.5 %, which is lower than it should be according to the inequality (1). These results are an inspiration to other simulation studies with a different choice of the calibration function parameters and possibly to modify the derived formula.

References

[1] Kubáček, L., Kubáčková, L.: *Statistics and metrology* (in Czech). Palacký University in Olomouc, Olomouc, 2000.

[2] Kenward, M. G., Roger, J. H.: Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53** (1997), 983–997.

[3] Myšková, K.: One-dimensional calibration with quadratic calibration function. *Forum Statisticum Slovacum* **2** (2011), 120–123.

[4] Osborne, C.: Statistical Calibration: A Review. *International Statistical Review* **59** (1991), 309–336.

[5] Rao, C. R.: *Linear methods of statistical induction and their applications.* Wiley, USA, 1973.

[6] Wimmer, G., Witkovský, V.: Scheffé-type confidence region for the calibration line. *Austrian Journal of Statistics* **35** (2006), 397–406.

Labour market frictions in a small open economy model of the Czech Republic

Daniel Němec[1]

Abstract. This contribution examines the impacts of introducing search and matching frictions in an open economy DSGE framework of the Czech economy. Model estimates should help to understand the driving forces behind important labour market variables: the wage bargaining power of unions, the match elasticity of unemployed, the efficiency of the matching process, separation rate and the flexibility of wages. Search and matching aspect provides satisfactory description of employment flows in the Czech economy. Moreover, the model of the open economy fits the business cycle features of the main economic variables better than the closed model.

Keywords: search and matching model, Bayesian estimation, DSGE model, small open economy.

JEL classification: C51, E24, J60
AMS classification: 91B40, 91B51

1 Introduction

The goal of my contribution is to reveal structural properties of the Czech labour market in the last twelve years and the main sources of unemployment dynamics. This contribution follows my previous research (see Němec [8]) in this area based on closed economy model and examines and compares the impacts of introducing search and matching frictions in an open economy DSGE framework of the Czech economy. For this purpose, I use a small open economy model with search and matching frictions incorporated into standard macroeconomic dynamic stochastic general equilibrium model (DSGE). This model is originally developed and estimated by Jakab and Kónya [7]. There are only few models in the literature which incorporate features to explicitly analyse labour market dynamics and search and matching frictions. In particular, Albertini et al. [1] investigate labour market dynamics in New Zealand economy by estimating a structural small open economy model enriched with standard search and matching frictions. A similar model augmented with different wage setting frameworks and frictions is estimated on Hungarian data by Jakab and Kónya [7]. Search and matching model is an important tool to model labour market dynamics. Model estimates should help to understand the driving forces behind important labour market variables: the wage bargaining power of unions, the match elasticity of unemployed, the efficiency of the matching process, separation rate and the flexibility of wages. One of the main questions of my contribution is how flexible is the Czech labour market. There is not an unique measure of the labour market flexibility but one can focus on some key features which might be connected with a flexible labour market. The labour market in the Czech Republic was influenced by the opening of markets which started in 1990. As Flek and Večerník [3] pointed out, the market reforms, trade and price liberalisation and the establishment of standard labour market institutions (aiming on improvement of labour mobility and flexibility) produced an inevitability of rising unemployment. Unlike other transition countries the rise of unemployment was delayed and unemployment rate hitted 10-years peak in 2004. Flek and Večerník [3] argues that the labour market alone was not fully responsible for this poor performance. Some obstacles (to better macroeconomic performance and job creation) were linked with a relatively weak supply-side flexibility of the Czech economy. These authors conclude that the Czech labour market loses its flexibility due to high reservation wage and due to the obstacles connected with the necessary layoffs. This conclusion is confirmed by Gottvald [4]. On the other hand, he pointed out that the diminishing flexibility in 90s was accompanied by the high probability of changing job (without an episode of unemployment). He

[1]Masaryk University, Faculty of Economics and Administration, Department od Economics, Lipová 41a, 602 00 Brno, Czech Republic, nemecd@econ.muni.cz

observed decreasing flows of workers among industries (i.e. low labour mobility). I am convinced that some of these issues may be confronted with the results of presented contribution.

2 Model

Small open economy DSGE model presented and estimated in this contribution is developed by Jakab and Kónya [7]. The model includes a non-Walrasian labour market with matching frictions and hiring costs. The model has two final good sectors, producing domestically sold and exported differentiated goods from a homogenous good. This homogenous good is produced using labour, capital, and imported intermediates. Search and matching frictions are incorporated only in this wholesales sector. Price rigidities are influencing only the final good producers. The wage bargaining and price setting decisions are thus separated. Imports are used as intermediate goods only. The model, its log-linearised form and steady-states equations are described in more detail in the original paper of Jakab and Kónya [7]. In this contribution, I will present only the main model concept and features which are necessary to understand the meaning and interpretations of model parameters. Moreover, some differences between the original paper and my specifications are emphasized.

The representative household maximizes intertemporal utility

$$\max_{c_t,i_t,b_t} E_0 \sum_{t=0}^{\infty} \beta^t \mathrm{e}^{\epsilon_t^c} \left[\frac{(c_t - hc_{t-1})^{1-\nu}}{1-\nu} - \chi_t n_t \right], \tag{1}$$

where c_t is consumption, n_t is employment rate, i_t is investment, b_t is bonds held by household expressed in local currency, χ_t is disutility from work (stochastic factor with expected value $\overline{\chi}$), ν is intertemporal elasticity of substitution and h represents habit persistence. Households constraints are:

$$c_t + i_t + \frac{b_t}{p_t R_t} = \frac{b_{t-1}}{p_t} + (1-u_t)w_t + u_t b_u + r_t^k K_{t-1} + d_t, \tag{2}$$

$$K_t = (1-\delta)K_{t-1} + \left[1 - \Phi\left(\frac{\mathrm{e}^{\epsilon_t^I} i_t}{i_{t-1}} \right) \right] i_t, \tag{3}$$

where u_t is unemployment rate, p_t is the consumer price index, w_t is the real wage rate, r_t^k is the (real) rental rate on capiral, K_t is the capital stock, d_t is lump sum net income from other sources. he investment adjustment cost $\Phi(\cdot)$ is increasing and convex with $\Phi(1) = \Phi'(1) = 0$ and $\Phi''(1) > 0$. Unemployment benefits b_u are financed by lump-sum taxes included in d_t. The wholesale sector produces a homogeneous product, using capital, imported intermediates and labour. The model assumes that each firm employs one worker. The Cobb-Douglas production function may be thus rewritten in a per-worker form:

$$y_t = \mathrm{e}^{a_t} k_t^{\alpha} y_{m,t}^{\alpha_z (1-\alpha)}, \tag{4}$$

where a_t is an exogenous shock, k_t is capital per-worker, $y_{m,t}$ is imported intermediates per-worker, α is the share of capital and α_z is the share of imported inputs intermediates. Demand for inputs is determined by the relative price of wholesale goods, $p_{w,t}$ and (exogenous) foreign currency price of the intermediate input, $p_{m,t}$. The labour market is subject to search and matching frictions. Each job may be destroyed with (exogenous) job destruction probability, ρ_t. Search frictions are captured by a standard Cobb-Douglas matching function:

$$m_t = \sigma_m \mathrm{e}^{\epsilon_{mf,t}} v_t^{\sigma} u_t^{1-\sigma}, \tag{5}$$

where m_t is the number of new matches, v_T is the number of open vacancies, u_t is the number of unemployed. Parameter $0 < \sigma < 1$ is a match elasticity of the unemployed and $\sigma_m \epsilon_{mf,t}$ is stochastic process measuring the efficiency of the matching process. Evolution of employment, n_t is given by

$$n_t = (1-\rho_t)n_{t-1} + m_t. \tag{6}$$

The model assumes that new matches are productive immediately, but the workers who loose their jobs have to wait one period before searching for a new job (Němec [8] assumes that it takes one period for new matches to be productive). Aggregate probability of filling a vacancy (job filling rate) may be defined as $q_t = m_t/v_t$, job finding rate $s_t = m_t/u_t$ and job market tightness is defined as $\theta_t = v_t/u_t$.

Wage setting processes are different for the wages of new hires and wages in existing jobs. Nominal wages w_t in existing jobs are bargained with a probability of $1-\gamma_w$, otherwise the wage remains at last period's wage. For new hires, the nominal wage is bargained with probability $(1-\nu_w)$. Unemployed workers receive unemployment benefit b_u which is important for determination of the wages. Negotiated wages w^* are set as a solution of the Nash bargaining problem, where parameter η represents bargaining power of the workers. The trajectory of the average wage is given by evolution of the newly set wages and by those wages which cannot be set optimally:

$$w_t = \frac{m_t}{n_t}\left[\nu_w w_{t-1} + (1-\nu_w)w_t^*\right] + \frac{(1-\rho_t)n_{t-1}}{n_t}\left[\gamma_w w_{t-1} + (1-\gamma_w)w_t^*\right]. \tag{7}$$

Job creation is fully derived in Jakob and Kónya [7]. The final good sector operates under the condition of an infinite number of monopolistically competing firms buying wholesale good (and differentiating it). Final goods, $y_t(i)$ are aggregated using the following (CES) utility function (index):

$$y_t = \left[\int_0^1 y_t(i)^{\frac{1}{1+\mu_t}} di\right]^{1+\mu_t}, \tag{8}$$

where μ_t is the time-varying markup parameter. The firms can reoptimize prices with probability $1-\gamma$. Remaining firms are able to increase their prices by indexation (parameter ν_p). Resulting Phillips curve is expressed by the means of perceived inflation. In this model, agents apply a real-time adaptive algorithm to identify the underlying inflation rate (learning algorithm is applied only to the domestic sector):

$$\hat{\pi}_t^p = \rho_\pi \hat{\pi}_{t-1}^p + gl(\hat{\pi}_t - \hat{\pi}_{t-1}^p) + \epsilon_{gl,t}, \tag{9}$$

where $\hat{\pi}_t$ is the observed inflation rate, $\hat{\pi}_t^p$ is the perceived inflation rate (expressed in log deviation from the steady state) and parameter gl influences the speed of learning. Jakab and Kónya [7] show that the price Phillips curve takes the same form as in a model without learning. In their specification, the inflation variable is the difference between the true and perceived inflation:

$$d\hat{\pi}_t = \hat{\pi}_t - \hat{\pi}_t^p. \tag{10}$$

Learning shock is an enhancement of the original model which allows a better identification of the model parameters. Moreover, due this enhancement, both the learning parameter and the autoregressive parameter are estimated consistently within the model identification procedure. The wholesale sector is composed into $n_{d,t}$ firms (and workers) producing domestically sold goods. These are used for consumption, investment and exogenous and unproductive government consumption (an AR(1) process):

$$n_{d,t} y_t = c_t + i_t + g_t. \tag{11}$$

Monetary policy is represented by a simple Taylor rule:

$$\hat{r}_t = \xi_r \hat{r}_{t-1} + (1-\xi_r)(\xi_\pi E_t \hat{\pi}_{t+1} + \xi_e \hat{e}_t) + \epsilon_t^m, \tag{12}$$

where monetary authority sets the interest rate (log deviation from the steady state) $\hat{r}_t$ as a reaction to the expected inflation gap, $E_t\hat{\pi}_{t+1}$, and the exchange rate gap, $\hat{e}_t$. The model describes a small open economy. In this case, a modified UIP (uncovered interest rate parity) holds, where the interest rate on home currency denominated foreign bonds is given by the constant world interest rate plus an endogenous risk premium:

$$\frac{e_t R_t}{E_t e_{t+1}} = \left[\frac{1}{\beta} + \psi(\mathrm{e}^{-(B_t-\overline{B})-1})\right]\mathrm{e}^{\epsilon_{uip,t}}, \tag{13}$$

where ψ is debt elasticity of financial risk premium (a function of the net foreign asset position B_t). Export demand equation in log-linearized form is $\hat{n}_{x,t}+\hat{y}_t = -\theta_x \hat{p}_{x,t} + \mathrm{e}^{\epsilon_{x,t}}$, where $\hat{}$ denotes the appropriate deviations fro steady states. Proportion of workers employed in the export sector is $n_{x,t} = n_t - n_{d,t}$, export price is $p_{x,t}$, θ_x determines the price elasticity of exports and $\epsilon_{x,t}$ is export demand shock. For further details see Jakab and Kónya [7]. All equations are log-linearized around the fixed steady state.

3 Estimation results and model evaluation

The data set for the Czech Republic used for estimation is from the first quarter 1996 to the fourth quarter 2011. The data comes from the OECD database. The observed variables are nominal short-term

Description	Parameter	Value	Source
Discount factor	β	0.99	Jakab and Kónya [7]
Steady-state share of capital	α	0.33	Herber and Němec [5]
Steady-state share of imported inputs	α_z	0.51	Herber and Němec [5]
Depreciation rate	δ	0.025	Herber and Němec [5]
Investments adjustment cost	$\Phi''(1)$	5	Herber and Němec [5]
Debt elasticity of financial risk premium	ψ	0.001	Jakab and Kónya [7]
AR parameter of government spending shock	ρ_g	0.6525	estimated outside the model
AR parameter of import price shock	ρ_{p_m}	0.8678	estimated outside the model
Average (steady state) mark-up in final goods	$\overline{\mu}$	0.20	Jakab and Kónya [7]
Average (steady-state) job finding rate	$\overline{s}$	0.0806	Hobijn and Sahin [6]
Average (steady-state) unemployment rate	$\overline{u}$	0.077	estimated outside the model
Average (steady-state) separation rate	$\overline{\rho}$	0.0067	$\overline{s} * \overline{u}/(1-\overline{u})$

Table 1 Calibrated parameter values and description

interest rate, real investments (intensive form – i.f.), real consumption (i.f.), real exports, real imports (i.f.), real government spending (i.f.), nominal wages, employment (based on unemployment level and rate), CPI inflation rate, import and export prices denominated in foreign currency, and the exchange rate.[1] All variables are expressed in logarithms and filtered using Hodrick-Prescott filter with the standard smoothing parameter ($\lambda = 1600$) for quarterly data. Parameters are estimated using Bayesian techniques. All computations have been performed using Dynare toolbox for Matlab (Adjemian et al. [2]). Table 1 reports calibrated model parameters. Table 2 describes prior densities of estimated parameter. The priors (and calibrations) are similar to those used by Jakab and Kónya [7] with regards to the patterns of the Czech economy (see Herber and Němec [5] and Němec [8]). The standard deviations are rather uninformative. The log-linearized model contains 37 equations, 11 AR(1) shocks and two uncorrelated shocks ($\hat{\mu}_t$, $\hat{\chi}_t$).

Table 3 presents[2] the posterior estimates of parameters and 90% highest posterior density intervals. It may be seen (in comparison with the Table 2) that most of the parameters are moved considerably from their prior means. The data seems to be strongly informative. The results for the open economy model are not too much different from those from the closed economy model (see Němec [8]). But, there are some interesting results which should be emphasized. The first surprising estimate is the bargaining power of workers, η. The mean value of this parameter is almost 0.4 with a 90% HPDI that is shifted away from the prior density. This implies that the firms can gain a lot of their entire surplus. The firms may be thus willing to create vacancies. Lower bargaining power of the workers is typical for the flexible labour markets which bring the wage dynamics to the line with productivity growth. The opposite results by Němec [8] are not confirmed. In the open economy framework, the separation rate parameter (its deviation from the calibrated steady-state value) is modelled as time-varying parameter. Smoothed trajectory of this variable provides a significant evidence of variability. Its steady-state value supports the view of less flexible Czech labour market with limited ability to destroy old and new matches. Low flexibility is meant to be associated with the restricted flows of the workers among industries. On the other hand, periods of the economic slowdown starting at the end of 2008 show that the separation rate is considerably higher – twice as big as the steady-state value. The estimate of parameter *bu* corresponds to a reasonable value of 0.30 for the Czech economy which is in accordance with the real unemployment benefits paid within the Czech social insurance system (similar estimate to the closed economy model). The posterior mean of the matching function parameter, σ, is in accordance with the common values in literature (see Němec [8] for further references). The trajectories of selected smoothed variables and shock innovations show (like the results for a closed economy model) a relative sharp decline in the

[1] Intensive form means that the variables are divided by the level of employment. Nominal wages (instead of the real wages) are used in the log-linearized version of the model. I used the following data sets: private final consumption, investments, government final consumption expenditure, exports and imports of goods and services (millions of national currency, chained volume, national reference year 2005, quarterly levels, s.a.); consumer price index (all items, 2005=100, s.a.); export and import price index (2000=100, s.a.); index of hourly earnings (manufacturing, 2005=100, s.a.); registered unemployment level and rate (s.a.); short-term interest rates (per cent per annum); exchange rate (USD/CZK).

[2] Due to the maximal allowed range of the contribution, all the figures (data, smoothed variables and shock, IRFs and shock decompositions) are a part of the corresponding conference presentation and may be obtained upon a request.

development of variable (probability of filling a vacancy) at the end of the year 2006. This evidence is in favour of theories which stressed the role of an obvious lack of employees in the Czech economy. This tendency was reverted as a result of the last global economic slowdown starting at the end of 2008. This downturn influenced a fall of the matching rates below their steady-state values. The starting recession has re-established the equilibrium on the labour market which may be documented by the trajectory of labour market tightness. The improvement of labour market institutions may be associated with the development of efficiency shock. No remarkable tendencies may be found on the Czech labour markets.

Description	Parameter	Density	Mean	Std. Dev.
Intertemporal elasticity in consumption	ν	Normal	1.50	0.40
Habit persistence	h	Beta	0.60	0.15
Indexation – final good price	ν_p	Beta	0.50	0.15
Indexation – export price	ν_x	Beta	0.50	0.15
Calvo parameter – final good sector	γ	Beta	0.50	0.15
Calvo parameter – exporters	γ_x	Beta	0.50	0.15
Calvo parameter – old wages	γ_w	Beta	0.50	0.15
Calvo parameter – new wages	ν_w	Beta	0.50	0.15
Bargaining power of the workers	η	Beta	0.50	0.15
Unemployment benefits	b_u	Beta	0.50	0.15
Steady state of disutility of labour	$\overline{\chi}$	Beta	0.20	0.05
Elasticity of matching	σ	Beta	0.70	0.20
Price elasticity of exports	θ_x	Beta	0.50	0.15
Taylor rule – exchange rate	ξ_e	Beta	0.30	0.10
Taylor rule – interest rate smoothing	ξ_r	Beta	0.50	0.15
Taylor rule – inflation	ξ_π	Normal	1.50	0.15
Adaptive learning parameter	gl	Beta	0.30	0.10
AR coefficients of shocks	$\rho_{\{a,mf,uip,x,I,m,\pi,\rho,c\}}$	Beta	0.50	0.15
Std. dev. of AR(1) shocks	$\sigma_{\epsilon_{\{c,I,a,mf,m,x,uip,\rho,gl\}}}$	Inv. Gamma	0.01	∞
Std. dev. of uncorrelated shocks	$\sigma_{\epsilon_{\{\chi,\mu\}}}$	Inv. Gamma	0.001	∞
Std. dev. of gov. consumption shock	$\sigma_{\{\epsilon_g\}}$	Inv. Gamma	0.03	0.01
Std. dev. of import prices shock	$\sigma_{\{\epsilon_{p_m}\}}$	Inv. Gamma	0.02	0.01

Table 2 Prior densities and description of estimated parameters

4 Conclusion

Unlike the results of Němec [8] or Albertini et al. [1], historical shock decomposition reveals the fact that the variation in unemployment and vacancies is not solely due to shocks to the labour market. There is a strong connection between the labour market and the rest of the economy. My estimates of Calvo parameters for old wages and new wages are $\gamma_w = 0.44$ and $\nu_w = 0.2$ respectively. Average duration of wage contracts are thus 1.8 quarters for the old hires and 1.25 quarters for the new matches. From this point of view, the wages are flexible. Adaptive learning mechanism and its parameters provide us with an evidence about a plausible montery policy in the Czech economy (the accomodation parameter gl is relatively low). The reaction function of the Czech National Bank is based on the parameters $\xi_r = 0.1$, $\xi_\pi = 1.4$ and $\xi_e = 0.41$. The smoothing parameter ξ_r is extremely low (comparing with the results by Herber and Němec [8]) and, on the other hand, exchange rate parameter ξ_e is unusually higher. There might be two explanations. Firstly, incorporating labour market features provide a new insight into the Taylor rule of the Czech economy. Secondly, omitting output gap from the Taylor rule may lead exactly to this kind of bias. Search and matching aspects provides satisfactory description of employment flows in the Czech economy. Historical shock decomposition may show that the model of the open economy fits the business cycle features of the main economic variables better than the closed model. Foreign demand plays a significant role in the development of the Czech economy and direct effects of labour market shocks on the economy and labour market are more obvious due to rich structure of the model.

	Posterior mean	90% HPDI			Posterior mean	90% HPDI	
ν	1.6752	1.4386	1.9717	h	0.9171	0.8547	0.9533
ν_p	0.2899	0.1938	0.4088	γ	0.6702	0.5769	0.7597
ν_x	0.3639	0.1762	0.5659	γ_x	0.5307	0.4805	0.5662
ν_w	0.1998	0.0964	0.2908	γ_w	0.4388	0.3918	0.4837
η	0.3967	0.3171	0.4607	bu	0.2965	0.2915	0.3015
$\overline{\chi}$	0.1347	0.0506	0.2205	σ	0.6384	0.5326	0.7618
θ_x	0.4492	0.3475	0.5468	gl	0.0925	0.0606	0.1245
ξ_r	0.0995	0.0435	0.1535	ξ_π	1.3944	1.3516	1.4230
ξ_e	0.4135	0.3923	0.4343	ρ_ρ	0.3366	0.2561	0.4141
ρ_c	0.4182	0.3593	0.4848	ρ_I	0.2460	0.1713	0.3174
ρ_a	0.3331	0.2717	0.3911	ρ_{mf}	0.5760	0.5401	0.6148
ρ_m	0.0718	0.0283	0.1149	ρ_x	0.7405	0.6825	0.8218
ρ_{uip}	0.1509	0.0938	0.2022	ρ_π	0.5710	0.4880	0.6508
σ_{ϵ_c}	0.4064	0.2396	0.5326	σ_{ϵ_I}	0.0711	0.0118	0.0157
σ_{ϵ_a}	0.0138	0.0118	0.0157	$\sigma_{\epsilon_{mf}}$	0.1279	0.0693	0.1810
σ_{ϵ_m}	0.2021	0.1709	0.3211	σ_{ϵ_x}	0.0497	0.0423	0.0567
$\sigma_{\epsilon_{uip}}$	0.2222	0.1824	0.2603	σ_{ϵ_g}	0.0302	0.0260	0.0351
$\sigma_{\epsilon_{pm}}$	0.0204	0.0177	0.0226	σ_{ϵ_ρ}	0.5723	0.4790	0.6584
σ_χ	0.0024	0.0002	0.0028	σ_μ	0.0811	0.0684	0.0930
$\sigma_{\epsilon_{gl}}$	0.0145	0.0050	0.0239				

Table 3 Parameter estimates

Acknowledgements

This work is supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/0780/2011.

References

[1] Albertini, J., Kamber, G., and Kirker, M.: *An estimated small open economy model with frictional unemployment*. Discussion Paper Series, DP2011/04, Reserve Bank of New Zealand, 2011.

[2] 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.

[3] Flek, V., and Večerník, J.: The Labour Market in the Czech Republic: Trends, Policies and Attitudes. In: *Anatomy of the Czech Labour Market: From Over-Employment to Under-Employment in Ten Years?* (Flek, V., ed.), CNB Working Paper Series 7-2004, 2004, 7–24.

[4] Gottvald, J.: Czech Labour Market Flows 1993–2003. In: *Anatomy of the Czech Labour Market: From Over-Employment to Under-Employment in Ten Years?* (Flek, V., ed.), CNB Working Paper Series 7-2004, 2004, 42–53.

[5] Herber, P., and Němec, D.: Investigating structural differences of the Czech economy: Does asymmetry of shocks matter? *Bulletin of the Czech Econometric Society* **19** (2012), 28–49.

[6] Hobijn, B., and Sahin, A.: *Job-Finding and separation Rates in the OECD*. Staff Report no. 298, Federal Reserve Bank of New York, 2007.

[7] 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.

[8] Němec, D.: Czech labour market through the lens of a search and matching DSGE model. In: *Proceedings of the Mathematical Methods in Economics 2011* (M. Dlouhý, and V. Skočdopolová, eds.), Praha, 2011, 504–510.

The impact of FDI on the host economy

Pavla Nikolovová[1]

Abstract. The foreign direct investment (FDI) and its impact on performance of domestic firms has been studied in many empirical papers, which, however, present rather ambiguous results. I argue that this is due to some limitations of prevalently used methodology, which does not separate the FDI spillover effects from the changes in competitive environment faced by domestic firms.

In my research, I propose a novel estimation strategy that allows me to disentangle FDI spillovers from the effects of competition changing in response to the entry of a foreign firm. I consider this issue on the industry level and I compare the effects of FDI to the impact of international trade on the domestic economy. My analysis covers the time period 2001 - 2007 and concerns both Western and Eastern European countries. My identification strategy leads me to confirm the presence of positive spillovers stemming from FDI.

Keywords: FDI, MNE, horizontal spillovers, international trade

JEL classification: F23
AMS classification: 91G70

1 Introduction

Foreign direct investment (FDI) can be characterized as an operation by which a multinational enterprise (MNE) acquires a substantial control over a domestic firm in the host economy. The volume of foreign direct investment in Central and Eastern European (CEE) countries has been increasing in the past twenty years, and it has generally been welcomed by domestic governments because the presence of a foreign firm is considered to have a strong potential to improve domestic economic conditions. Instead of confirming this common expectation, however, empirical studies draw rather ambiguous conclusions as to whether the consequences of FDI are indeed as significant and as positive as it is believed.

My paper focuses on the impact of FDI on domestic firms within the same industry, and it complements the research that has been done in this field, which is very extensive but which still leaves many questions without clear and definitive answers[1]. The major question I address here is how to disentangle the "competition effect" and the "spillover effect", which, as Kosová [4] explains, both take place when domestic firms in an industry have to face a highly efficient MNE entering the market. At first, the MNE increases the competition within the industry, making some of the domestic firms leave the market (the competition effect). Only then there can be positive spillover effects stemming from the interaction between the MNE and the surviving domestic firms: these firms have a positive example which they can follow, they can copy the technologies, they can hire workers or managers that have had experience working for the MNE, and so on.

The problem of existing empirical literature is that authors usually measure the negative competition effect and the positive spillover effect simultaneously, without being able to distinguish exactly what is the role of each of them. As a result, the overall effect can be misinterpreted as being positive, negative or insignificant, depending on which effect is offsetting the other. The purpose of my paper is to propose an identification strategy that would allow me to separate the competition effect from the spillover effects within a reduced form model.

The motivation for my estimation strategy stems from a theoretical model proposed by Helpman et al. [3], who study under which conditions a firm decides to export or to invest abroad. This decision is known as a "proximity-concentration tradeoff": when a firm wants to serve a foreign market, it can either undertake an investment in the country in question (by buying a local enterprise or by founding

[1]CERGE-EI, Politických vězňů 7, Praha 1, Czech Republic, e-mail: pavla.nikolovova@cerge-ei.cz
[1]For a detailed survey, see [2].

a new one) or it can export. The existence of this tradeoff provides me with the argument that foreign competition in the domestic market can be of two sources. It can come from foreign firms that are located abroad through imports, or from foreign firms that are implanted in the country as MNEs through FDI.

On the other hand, if the foreign firm is located abroad, there is no scope for technological spillovers which should occur only if domestic firms are in close contact with the MNEs - this argument being supported by [1]. Therefore, the idea of my estimation strategy is to compare the impact of foreign presence given both by imports from abroad and by MNEs operating in the country. The difference between these two allows me to see if the physical presence of MNEs in a country has some advantages for the host economy compared to the situation when the country is just exposed to international trade, and hence to filter out a potential spillover effect of FDI.

2 Data description

My analysis covers the time period 2001 - 2007 and focuses on European countries, which are considered to be either *Western* or *Eastern* countries. The Western countries are the countries of EU15 (Luxembourg being joint with Belgium) plus Iceland, Norway and Switzerland. The Eastern countries are the countries that joined the EU in 2004 or 2007. The analysis is performed on both groups separately to see the differences between fully developed countries and those who had just undergo the transition period (CEE countries). Special focus is on the countries of the Visegrad group (Czech Republic, Slovakia, Poland and Hungary), for which the analysis is provided apart.

I use the AMADEUS database to obtain information about firms operating in the chosen countries: their performance, their financial and organizational characteristics, their ownership structure (especially if they are domestic or foreign) and their industry classification given by the three-digits NACE code. I link this database with information from UN COMTRADE data about international trade, which covers international exports and imports between the selected countries and their trade partners in the studied time period, disaggregated to the four- and five-digits SITC level.

Unlike in other papers concerning the issue of FDI, my analysis is performed on industry level. The motivation for this approach is simple: first, I am not interested in the impact of the foreign presence on particular firms, but rather on the average efficiency of domestic firms in an industry, and second, both FDI and trade are sectoral variables and so there is no individual source of variation of these variables on firm level. The aggregation on industry level is performed by taking averages within industry, weighted by the share of domestic ownership in firms (as I am analyzing the performance of domestic firms only), and the level of aggregation is the four- and five-digits SITC level, leading to approximatively 250 000 observations structured as an unbalanced panel of industries in the above mentioned countries.

3 Econometric specification

Following the seminal paper [5], I estimate an augmented production function: I choose sales ($Sales$) to proxy the output variable, and tangible fixed assets ($Assets$) and number of employees ($Employment$) to proxy the factor inputs of capital and labor, with all variables being in logarithms; moreover, I include in my specification the lagged values of output to account for the imperfect allocation of factor inputs. I augment this production function by variables indicating the foreign presence: I use the variable $Imports$ to account for the foreign presence given by import flows and the variable FDI to account for the presence of MNEs.

I define imports as the volume of imported goods normalized by the size of the industry:

$$Imports_{it} = \frac{ImpVol_{it}}{\sum_{j=1}^{N_{it}} Sales_{ijt}} \quad ,$$

where t is time, N_i is the number of firms in industry i, $Sales_{ij}$ are the sales of the j-th firm in industry i and $ImpVol$ is the volume of imported goods in industry i.

To define the variable FDI, I have to take into account the issue of timing. I do not really expect the spillover effect, if there is such, to take place instantaneously. In my opinion, even if domestic firms could benefit from the presence of FDI, they would need some time to accommodate and to incorporate

possible technological improvements into their production. Therefore, I define the foreign presence given by FDI as the ratio of the sales of foreign owned firms in a given industry over the sales of all firms operating in that industry, but I take into account only firms that were already foreign owned in the previous year:

$$FDI_{it} = \frac{\sum_{j=1}^{N_{it}} f_{ijt-1} Sales_{ijt}}{\sum_{j=1}^{N_i t} Sales_{ijt}} \quad ,$$

where f_{ij} is the share of foreign owners in the j-th firm in industry i and otherwise the notation is the same as for the definition of $Imports$.

From these definitions, we can see that whereas by construction, the variable FDI is from the interval $[0, 1]$, the variable $Imports$ can have any positive value. The reasons are that first, there is nothing that prevents the imports to be larger than domestic production and second, whereas from the UN COMTRADE, I have the complete information about international trade, from the AMADEUS database, I have only a representative (even though very large) sample of firms and so I do not capture the whole domestic production. This implies that the two variables are measured in very different units. As it is usual in such cases, I decided to standardize both variables by dividing them by their standard deviations to get them on a comparable scale.

Using the above described variables, my econometric specification is

$$\begin{aligned} \ln(Sales_{it}) \quad = \quad & \beta_0 + \beta_1 \ln(Sales_{it-1}) + \beta_2 \ln(Assets_{it}) + \beta_3 \ln(Employment_{it}) + \\ & + \delta_{FDI} FDI_{it} + \delta_{Imports} Imports_{it} + \gamma_i + \gamma_t + u_{it} \quad . \end{aligned} \tag{1}$$

I include time and industry fixed effects (the industry being in fact an industry-state unit, because I aggregate over firms in industries only within countries, not across).

Every estimation is run twice: first on the whole sample of industries in the given geographical region, and second on industries that are not oriented to exporting. The estimation on the whole sample is presented basically for the sake of completeness of my analysis. My identification strategy, which is based on filtering out the competition effect, can work only when we talk about the competition in the domestic market, because I compare imported goods (which are obviously sold only in the domestic market) to sales of firms operating in the industry. If a significant part of the production of domestic firms goes for export, then my identification strategy cannot really work.

In reality, most of the industries have both import and export flows, because they are industries with differentiated products. Hence, I cannot really find an industry that would be purely import oriented and as a consequence, my identification strategy is not flawless. However, I can at least focus on industries that are less export oriented than others, which is why I run for each geographical region a second estimation only on a subsample of industries where the exports (normalized by total sales) are below the median of the whole sample. This is the estimation that I focus on when evaluating my research hypotheses, presented in the following section.

4 Hypotheses

The literature on spillover effects claims that if these are present, the coefficient δ_{FDI} in (1) should be positive; however, it is often found insignificant or negative. I argued throughout this paper that this might be because the variable FDI influences the output in two opposite ways: by inducing the negative competition effect and the positive spillover effect at the same time.

To account for the foreign competition, I introduce in the model the variable $Imports$, which should also represent the negative competition effect but no positive spillover effect. To verify this, I test if the coefficient $\delta_{Imports}$ in (1) is negative:

Hypothesis 1:

$$H_0 : \ \delta_{Imports} \geq 0 \quad \text{vs} \quad H_A : \ \delta_{Imports} < 0 \quad .$$

Further, I compare the coefficients δ_{FDI} and $\delta_{Imports}$ in (1), and if their difference is positive, I can conclude that there is a positive spillover effect present, which outweighs, at least partially, the negative competition effect:

Hypothesis 2:

$$H_0: \ \delta_{FDI} - \delta_{Imports} \leq 0 \quad \text{vs} \quad H_A: \ \delta_{FDI} - \delta_{Imports} > 0 \ .$$

Hence, the rejection of H_0 of the first hypothesis justifies my identification strategy and the rejection of H_0 of the second hypothesis proves the presence of positive spillover effects of FDI.

5 Results

The results of regression (1) are presented in Table 1. The results for the estimation over the whole sample can be found in the first three columns (for Western, Eastern and Visegrad group countries), the results for the subsample of non export oriented industries are in the the last three columns. Several observations can be made based on these.

	All industries			Non export oriented		
	West	East	Visegrad	West	East	Visegrad
FDI	-0.138***	-0.066***	-0.062***	-0.103***	-0.058***	-0.058***
	(0.004)	(0.003)	(0.004)	(0.005)	(0.004)	(0.007)
Imports	-0.100***	-0.051***	-0.038***	-0.148***	-0.126***	-0.246***
	(0.004)	(0.005)	(0.006)	(0.032)	(0.031)	(0.088)
Lagged Sales	-0.015***	-0.014***	-0.008**	0.012***	0.008*	0.009*
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.005)
Tangible fixed assets	0.377***	0.498***	0.565***	0.433***	0.520***	0.694***
	(0.007)	(0.005)	(0.007)	(0.010)	(0.007)	(0.010)
Employment	0.330***	0.299***	0.209***	0.285***	0.311***	0.133***
	(0.006)	(0.007)	(0.007)	(0.010)	(0.010)	(0.010)
Year effects	Yes	Yes	Yes	Yes	Yes	Yes
Industry-country effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.489	0.712	0.737	0.517	0.744	0.797
Observations	173480	81392	36803	86737	38691	18331

Clustered (on industry level) standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1: Estimation results

First, it has to be said that in all subsamples, the effect of FDI (when measured solely by the coefficient on this variable) is estimated as negative. This is in line with the results of many other papers concerning this topic and also with the metaanalysis proposed in [2]. Hence, if my conclusions differ from those already published, it is not because of the construction of my dataset nor my variables, but just because of my identification strategy.

Second, in all subsamples, the effect of $Imports$ is negative and significant. I can thus reject the null of *Hypothesis 1* and conclude that imports really induce a negative competition effect on domestic firms, supporting thus the assumption on which my estimation is based.

Third, the results differ in the estimation performed over the whole dataset as compared to the estimation over the subsample of non export oriented firms. In line with my expectations, in the latter one, the negative effect of imports is more pronounced: in this subsample, domestic firms serve the domestic market and compete with imported goods. This proves that to answer my research question about the presence of positive spillovers, I should focus only on the subgroup of non export oriented firms, because it is relevant to my identification strategy.

Fourth, if we focus only on this subgroup, we can see that the coefficient on FDI is consistently less negative than the coefficient on $Imports$, which indicates that the null of *Hypothesis 2* is likely to be rejected. I test this hypothesis formally using a one-sided t-test of the difference of the two coefficients. The results of this test are presented in Table 2 separately for the three regions, and they confirm that I can reject H_0 at 95% confidence level for Eastern countries and countries of the Visegrad group, whereas for Western countries, the statistical significance is a little bit less strong (but still valid at 90% confidence level).

	Western countries	Eastern countries	Visegrad group
t-statistic	1.407	2.152	2.121
p-value	0.080	0.016	0.017

Note: p-values of asymptotic one-sided test (standard normal distribution)

Table 2: Hypothesis testing

To sum up, I can reject the null hypothesis that the effect of FDI is more negative than the effect of imports: I find the difference of these two effects to be positive. This result supports the theory that there might be positive spillover effects stemming from FDI, and thus answers my main research question, but it merits to be commented a little further.

It is interesting to compare the estimation results for the three geographical regions. If we define the spillover effect as the difference between the coefficients on FDI and on $Imports$, we see it is the largest for countries of the Visegrad group and relatively smaller for Eastern countries as a whole and for Western countries. This result can be interpreted in line with other papers analyzing the effect of FDI: it is hypothesized that to internalize the spillover effect, domestic companies should not be too inferior in terms of efficiency to MNEs, because when the efficiency gap is too wide, domestic companies are not able to "catch up". Hence, the spillover effect is a U-shaped function of domestic firms' efficiency: if domestic firms' efficiency is very small compared to MNEs, the spillover effect is weak because of the inability to internalize; if domestic firms' efficiency is similar to the efficiency of MNEs, the spillover effect is also weak because there is not too much scope for improvement; if the gap between domestic firms and MNEs is significant but moderate, the spillover effect is the strongest[2].

If we assume that domestic firms in Western countries are the closest to MNEs in terms of efficiency, we should not be surprised that there is not a very significant spillover effect present - there is not too much to learn from the point of view of domestic firms. Further, a closer look on the data tells us that firms of the Visegrad group are closer in their characteristics to Western firms than the mean of Eastern firms is, which signals that within the group of Eastern countries, the countries of the Visegrad group are rather above the average, and domestic firms there are then also closer to MNEs in terms of efficiency, even though the gap is still significant. This observation together with the above presented theory of the U-shaped effect could explain the differences I found among regions.

6 Robustness check

It has to be admitted that the validity of the presented results depends on how well the chosen variables proxy the control variables of the theoretical model of production function, especially labor and capital. In my estimation, I choose total fixed assets and number of employees, because these are often used in the stream of literature to which I am relating my analysis. However, as a robustness check, I also repeated the estimation using working capital as a measure of capital used for production and staff costs as a measure of labor. The results, which are available upon request, were not significantly different from those presented here.

7 Conclusion

In this paper, I contributed to the literature concerning the impact of FDI on the host economy by presenting a new identification strategy for the horizontal spillover effect. I explained why this effect is

[2]For more details, see [6].

not correctly identified in papers that take into account only the presence of firms with foreign owners in the domestic market. I pointed out that the positive spillovers might be outweighed by a negative competition effect if the competition environment is not controlled for. My strategy for identification of spillovers is to compare the effect induced by foreign firms that import in the domestic market with the effect induced by foreign firms that actually operate in the domestic economy: the difference between these two effects should be attributed to potential spillovers. I performed the analysis on a large panel of industries in European countries in the period 2001-2007.

The key contribution of my paper lies in the comparison of the two sources of foreign presence which enables me to properly isolate the spillover effect and to confirm its positive impact on the performance of domestic firms. This finding is especially pertinent for the countries of the Visegrad group.

Acknowledgements

I would like to express my gratitude to my advisor prof. Jan Hanousek for his help and encouragement without which this paper would never have been completed. I would also like to thank Mário Vozár for his valuable observations and comments.

This work was supported by the GAČR grant No. 403/12/0080, by the grant SVV-2011-263 80, by the GAUK grant 43010 and by the CEFRES / Komerční banka scholarship 2010/2011.

References

[1] Ethier, W.: The Multinational Firm. *Quarterly Journal of Economics* **10(4)** (1986), 805–834.

[2] Hanousek, J., Kocenda, M. and Maurel, M.: Direct and Indirect Effects of FDI in Emerging European Markets: A Survey and Meta-analysis. William Davidson Institute Working Papers Series, No. 976. (2010).

[3] Helpman, E., Melitz, M. J. and Yeaple, S. R.: Export Versus FDI with Heterogeneous Firms. *American Economic Review* **94(1)** (2004), 300–316.

[4] Kosová, R.: Do Foreign Firms Crowd Out Domestic Firms? Evidence form the Czech Republic. *Review of Economics and Statistics* **92(4)** (2010), 861–881.

[5] Nickell, S. J.: Competition and Corporate Performance. *Journal of Political Economy* **104(4)** (1996), 724–746.

[6] Smeets, R.: Collecting the Pieces of the FDI Knowledge Spillovers Puzzle. *Research Observer, International Bank for Reconstruction and Development* (2008).

Modelling corporate bond rating with the use of market–based indicators

Martina Novotná[1]

Abstract. The paper presents an estimation of corporate bond rating models based on both financial and market company indicators. The analysis will be carried out for companies in the oil and gas industry having a rating assessment from Moody's rating agency. The paper aims to offer a more detailed understanding of the relationship between company market indicators and bond rating classification. Bond rating models will be estimated by multivariate statistical methods such as discriminant analysis and logistic regression. The contribution of the paper is to identify variables with a significant impact on corporate bond rating in the selected industry. The models being derived allow classifying bond rating of companies with relatively high accuracy, even when a limited set of input variables is considered. The practical use of models lies in the area of management decision process and managing credit risk.

Keywords: Discriminant analysis, estimation, logistic regression, rating model.

JEL Classification: C35
AMS Classification: 62H12

1 Introduction

The aim of this study is to examine and quantify relationships among rating and other relevant data. The primary question is whether financial and market variables affect bond rating. If the answer is positive, the next question is what the nature of their relationship is. The study is based on cross-sectional data of a variety of companies from oil and gas industry, mostly from the United States. The whole sample covers 155 companies with Moody's rating; for the purposes of validation, it was split into two sub-samples. Experimental sample (approximately 75 %) will be used for model estimation and the remaining part (test sample) will be used for validation of models. Two methods will be used to estimate bond rating models, multinomial logistic regression and multivariate discriminant analysis. The next paragraph describes the methodology; results and classification ability of models will be assessed in the following chapters of this paper.

2 Overview of the methodology

This chapter is focused on a brief overview of two methods that will be used to estimate bond rating models, discriminant analysis and logistic regression analysis. The latter method became one of the most used methods to estimate bond rating or default prediction, see for example Altman, Sabato and Wilson [2], Waagepetersen [9], or Westgaard and Wijst [10]. An alternative and traditional approach to predict bond ratings is discriminant analysis introduced for example by Pinches and Mingo [6], Ang and Patel [3], or Altman and Eisenbeis [1].

Discriminant analysis

Discriminant analysis is a common statistical method used for separation of groups, and hence a suitable method for bond rating modelling. Discriminant functions are linear combinations of variables that best separate groups, for example the k groups of multivariate observations. In the following part of this subchapter, the explanations and definitions were taken from Rencher (2002, p. 277 – 286) [7].

For k groups with n_i observations in the ith group, we transform each observation vector y_{ij} to obtain $z_{ij} = a'y_{ij}$, i = 1, 2,..., k; j = 1, 2,...,ni, and find the means $\overline{z}_i = a'\overline{y}_i$, where $\overline{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i$. We seek the vector a that maximally separates $\overline{z}_1, \overline{z}_2, \ldots \overline{z}_k$. The separation criterion among $\overline{z_1}$, $\overline{z_2}$ ..., $\overline{z_k}$ can be expressed in term of matrices,

$$\lambda = \frac{a'Ha}{a'Ea} \tag{1}$$

[1] Technical University of Ostrava, Department of Finance, Sokolská třída 33, Ostrava, martina.novotna@vsb.cz.

where matrix H has a between sum of squares on the diagonal for each of the p variables, and matrix E has a within sum of squares for each variable on the diagonal. Another expression of the separation criterion is

$$\lambda = \frac{\text{SSH (z)}}{\text{SSE (z)}} \tag{2}$$

where SSH (z) and SSE (z) are the between and within sums of squares for z. The main task of the discriminant analysis is to find a set of weights (a values) for the outcome variables to determine a linear composite:

$$Z = a_1Y_1 + a_2Y_2 + \cdots + a_pY_p \tag{3}$$

so that the ratio (2) is maximized. The discriminant analysis follows by assessing the relative contribution of the $y's$ to separation of several groups and testing the significance of a subset of the discriminant function coefficients. The discriminant criterion (1) is maximized by λ_1, the largest eigenvalue of $E^{-1}H$; the remaining eigenvalues correspond to other discriminant dimensions. The test of significance is usually based on the Wilks' lambda, Λ, the most widely used criterion. The test statistic at the mth step is

$$\Lambda_m = \prod_{i=m}^{s} \frac{1}{1+\lambda_i}, \tag{4}$$

which is distributed as $\Lambda_{p-m+1,k-m,N-m+1}$. The statistic,

$$V_m = -\left[N - 1 - \frac{1}{2}(p+k)\right] \ln\Lambda_m = \left[N - 1 - \frac{1}{2}(p+k)\right] \sum_{i=m}^{s} \ln(1+\lambda_i), \tag{5}$$

has an approximate χ^2-distribution with (p-m+1)(k-m) degrees of freedom.

Logistic regression analysis

Hosmer and Lemeshow (2000, p. 31) [5] define the multiple logistic regression model as follows. A collection of p independent variables is denoted by the vector $\mathbf{x'} = (x_1, x_2, \ldots, x_p)$, assuming that at each of these variables is at least interval scale. The conditional probability that the outcome is present is denoted by $P(Y = 1|\mathbf{x}) = \pi(\mathbf{x})$. Then, the logit of the multiple logistic regression model is given by the equation,

$$g(\mathbf{x}) = \beta_0 + \beta_1.x_1 + \beta_2.x_2 + \cdots + \beta_p.x_p \tag{6}$$

and the logistic regression model is expressed by the following formula,

$$\pi(\mathbf{x}) = \frac{e^{g(\mathbf{x})}}{1 + e^{g(\mathbf{x})}}. \tag{7}$$

Based on De Laurentis (2010, p. 54 – 55) [4], the g (.) function (6) is known as a link function, which links variables xj and their coefficients βj with the expected value E(Yi) = πi of the ith observation of Y. The link function can be defined as the logarithm of the ratio between the probability of event (e.g. default) and the probability of non-event (e.g. non-default). This ratio is known as "odds" and can be formulated as follows:

$$g(\mathbf{x}) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1.x_1 + \beta_2.x_2 + \cdots + \beta_p.x_p \ . \tag{8}$$

The logit function associates the expected value of the dependent variable to a linear combination of the independent variables. The relationship between independent variables and the probability of default π is nonlinear, while the relationship between logit (π)and independent variables is linear.

Consider we have a sample of n independent observations $(\mathbf{x_i}, y_i)$, i=1,2,...*n*. Fitting the model requires to estimate vector $\boldsymbol{\beta}' = (\beta_0, \beta_1, \ldots, \beta_p)$ by the maximum likelihood method. The likelihood function can be described by the following formula, according to Hosmer and Lemeshow (2000, p. 8) [5]:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \pi\,(x_i)^{y_i}[1 - \pi(x_i)^{1-y_i}], \tag{9}$$

where $\pi(\mathbf{x})$ is defined as (7).

Assume $\widehat{\boldsymbol{\beta}}$ is the solution to the likelihood equations, then the fitted values for the multiple regression model are $\hat{\pi}(\mathbf{x_i})$, the value of the expression (8) computed using $\widehat{\boldsymbol{\beta}}$ and $\mathbf{x_i}$.

The multinomial logistic regression allows predicting membership of more than two categories. In this case, it breaks the outcome variable down into series of comparisons between two categories. In the analysis below, bond rating is a dependent variable, which has four possible outcomes. The existence of four categories requires three logit functions and determination of the baseline category, which is then compared with other logits.

3 Sample description

Companies with Moody's rating assessment have been considered in this study, the relevant data come from Moody's official websites[2], companies' annual reports and Yahoo! Finance websites[3] of business finance, stock market, quotes and news. After checking the data and adjustment for companies without all data available, the final sample consists of 155 companies. For the reasons of calculations[4], original rating categories have been re-coded as presented in the Table 1. The first three highest categories have been merged together because of a small number of representative companies, which could negatively affect results and stability of models.

Rating category	Rating code	Number of cases	Marginal percentage
Aaa, Aa, A	4	21	13.5 %
Baa	3	59	38.1 %
Ba	2	30	19.4 %
B	1	45	29.0 %
Total		**155**	**100%**

Table 1 Sample structure

The selection of independent variables should be thoroughly considered, because the set of input variables can substantially affect results, specifically predictive ability and stability of final models. The analysts usually stand on their previous results, experience and other research studies. Basically, most models are estimated based on financial statements of companies. Many studies prove that relatively simple rating models containing basic financial indicators provide good classification ability and can be used as a tool to assign a rating classification.

There are many possible **financial indicators** that can be used in the analysis. The selected indicators should reflect profitability, activity, liquidity and capital structure of companies and all of them should have a relationship with rating. To use some variables in the analysis, main assumptions should be met. First, the variables should have a normal distribution; secondly, multicolinearity should be avoided. In this study, the following financial variables are considered initially:

1. Total assets (*TA*);
2. Equity to total assets ratio (*Equity_to_TA*);
3. Long term debt to total assets ratio (*LTD_to_TA*);
4. Short term debt to total assets ratio (*STD_to_TA*);
5. Return on assets (*ROA*);
6. Return on equity (*ROE*);
7. Return on capital employed (*ROCE*);
8. Interest coverage (*Int_cov*);
9. Current ratio (*Curr_ratio*);
10. Total assets days outstanding (*Days_TA*).

The relationship between each variable and rating should have an economic rationale. For example, we can assume that the higher the size of total assets, the higher the protection of company's creditors, and the higher the rating category. Some variables had to be transformed to approach a normal distribution, such as TA (*LogTA*), Int_cov (*LogInt_cov*), Curr_ratio (*LogCurr_ratio*), Days_TA (*LogDays_TA*).

The main task of this study is to investigate the relationship among rating and selected **market-based variables** such as *beta, earnings per share, enterprise value and market capitalisation*. The paper should answer the question if these market indicators are related to rating. If so, the next step would be to investigate this relationship and use the market indicators to estimate bond rating models. To approach a normal distribution, some of these variables have been transformed (*LogMarketCap*, *LogEV*, *LogBeta*).

4 Bond rating models

Discriminant analysis (DA) and multinomial logistic regression (MLR) will be carried out to identify variables most relevant to rating classification. Two approaches will be used, the method in which all independent variables are included in the model (full), and stepwise method (step), which aims to include only the most significant variables in the model.

[2] http://www.moodys.com/
[3] http://finance.yahoo.com/
[4] PASW Statistics 18

4.1 Estimation of models

First, bond rating models will be estimated from financial data only. Then, only market-based data will be used and finally, results will be compared and a combination of both previous approaches will be applied.

Estimation of models with financial variables

The original set of independent variables was modified and some financial variables (3, 7, 10) were removed for the reasons of high correlations with other variables. The results (Table 2) show that by using only one financial variable (LogInt_cov), it is able to achieve similar classification ability as in the case of a model with seven variables. Overall, multinomial logistic regression provides better classification ability than discriminant analysis.

Model	Approach	Number of input variables	Number of variables in the model	Variables included	Classification ability
(A)	DA Full	7	7	All	46.8 %
(B)	DA Step	7	1	LogInt_cov	44.4 %
(C)	MLR Full	7	7	All	56.8 %
(D)	MLR Step	7	3	Equity_TA, LogInt_cov, LogCurr_ratio	52.3 %

Table 2 Models with financial variables

Estimation of models with market-based variables

Analogically to the previous case, both discriminant analysis and multinomial logistic regression were used to estimate the models and find the most significant indicators for classification. The results (Table 3) show that considering companies' market data only, model with much better classification ability can be obtained. The most relevant variables are EPS and LogEV.

Model	Approach	Number of input variables	Number of variables in the model	Variables included	Classification ability
(E)	DA Full	4	4	All	64.3 %
(F)	DA Step	4	2	EPS, LogEV	62.7 %
(G)	MLR Full	4	4	All	79.7 %
(H)	MLR Step	4	1	LogEV	70.9%

Table 3 Models with market-based variables

Combination of financial and market-based variables

When all the independent variables enter the analysis, the overall classification ability gently rises, especially in the case of MLR. By using all 11 variables, classification ability of 89.6 % can be achieved. By applying stepwise methods, the final models contain only two indicators, LogCurr_ratio and LogMarketCap (Table 4).

Model	Approach	Number of input variables	Number of variables in the model	Variables included	Classification ability
(I)	DA Full	11	11	All	62.7 %
(J)	DA Step	11	2	LogCurr_ratio LogMarket_Cap	66.7 %
(K)	MLR Full	11	11	All	89.6 %
(L)	MLR Step	11	2	LogCurr_ratio LogMarket_Cap	76.1 %

Table 4 Combination of financial and market-based variables

Modifications and adjustments

Based on the results above, it is evident that some variables contribute to classification more than the others. The final models would stand on the previous results and use only four predictors with the most significant discriminating power on rating, such as *LogInt_cov, EPS, LogEV* and *LogMarketCap*. Classification ability of the adjusted models is in Table 5.

Model	Approach	Number of input variables	Number of variables in the model	Variables included	Classification ability
(M)	DA Full	4	4	All	64.3 %
(N)	MLR Full	4	4	All	76.3 %

Table 5 Modification of models

The overall results suggest that market indicators contribute to the discrimination more than financial ratios. By adding market data to the original set of financial ratios, the total classification ability of models increases. Both methods, the discriminant analysis and multinomial logistic regression, provide similar results, however models estimated by MLR achieve higher classification ability. The best model from this point of view was estimated by MLR and uses all 11 financial and market variables (Model K). Good classification results are then achieved by MLR models using either 4 market indicators (Model G), or 4 combined variables (Model N). The overall results are surprising because they suggest that earnings per share, enterprise value, market capitalization and beta can give a good signal of a bond investment quality.

4.2 Verification and validation

Based on the criterion of classification ability on the original sample, the following three models, (G), (K) and (N) will be examined in more detail. All these models have been estimated by multinomial logistic regression, which allows simpler comparing of results and overall fit of models.

Criterion	Model G	Model K		Model N
Number of predictors:	4	11		4
Predictors included:	EPS	LogTA	LogCurr_ratio	LogInt_cov
	LogMarketCap	Equity_to_TA***	EPS	EPS
(*Likelihood ratio tests of parameters*)	LogEV***	STD_to_TA**	LogMarketCap	LogMarketCap
	LogBeta***	ROA*	LogEV**	LogEV*
		ROE*	LogBeta**	
		LogInt_cov		
Model fitting:				
Chi-Square	133,286*** (df=12)	150,528*** (df=33)		132,122*** (df=12)
Goodness-of-Fit:				
Pearson	184.708 (df=219)	28.304 (df=165)		157.824 (df=264)
Deviance	78.954 (df=219)	30.030 (df=165)		114.580 (df=264)
Measures of R^2:				
Cox and Snell	0.815	0.758		0.894
Nagelkerke	0.875	0.816		0.959

****p<.001, ** p<.01, * p<.05*

Table 6 Verification

To assess the fit of models, we use a log-likelihood statistic, which is based on summing the probabilities associated with the predicted and outcome variables, Tabachnik and Fidell (2007, p. 446) [8]. The statistic indicates how much unexplained information there is after the model has been fitted. The larger the value, the more unexplained observations there are. The chi-square test tests the decrease in unexplained variance from the baseline model to the final model. All the final models explain a significant amount of the original variability, so they better fit than the original model. The next test tests whether the models predicted values are significantly different from the observed ones. If the statistics (Pearson and Deviance) are not significant, than predicted and observed values are not different, and the model is a good fit. All three models are a good fit based on this test. The significance of predictors to the models was assessed by the likelihood ratio tests. In all models, variable LogEV

has a significant main effect on rating category classification; it is even the only significant predictor in Model N. Due to a large number of derived models in this study, parameter estimates and odds ratios are not included in this paper, however they can be provided on demand.

The three selected models (G, K, N) were used to predict bond rating of companies other than that used for estimation of models. As the test sample covers only 25 companies, results of the validation will not likely be accurate and can be misleading.

Model	Correct classification 4-rating	Correct classification 2-rating
Model G	16 %	64 %
Model K	28 %	60 %
Model N	32 %	84 %

Table 7 Validation

As expected, the ratio of correctly classified companies is very low, which is likely the result of relatively small control sample. When classifying into four rating groups, all three models give bad results. However, all models contribute significantly to the classification in case of just two rating groups, investment and speculative category. Validation proved that that the Model N provides very accurate predictions.

5 Conclusion

The overall results suggest that market-based indicators contribute to the discrimination more than financial ratios. By adding market-based data to the original set of financial ratios, the total classification ability of models increases. Both methods, discriminant analysis and multinomial logistic regression, provide similar results, however models estimated by MLR achieve higher classification ability. The best model from this point of view was estimated by MLR and uses all 11 financial and market variables (Model K). Good classification results are then achieved by MLR models using either 4 market-based indicators (Model G), or 4 combined variables (Model N). The overall results are surprising because they suggest that earnings per share, enterprise value, market capitalisation and beta can indicate the bond rating category of companies relatively accurately. The most significant variable for bond rating prediction is LogEV. Thus, the enterprise value can be a very important indicator of creditworthiness and can give a good signal of a bond investment quality.

Acknowledgements

The paper is based on research activities sponsored through SGS research project No. SP2012/19. The support is greatly acknowledged.

References

[1] Altman, E. and Eisenbeis, R.: Financial applications of discriminant analysis: A clarification. *Journal of Financial and Quantitative Analysis* **13** (1978), 185 – 195.
[2] Altman, E.I., Sabato, G. and Wilson, N.: The value of non-financial information in small and medium-sized enterprise risk management. *The Journal of Credit Risk* **6** (2010), 95 – 127.
[3] Ang. J.S. and. Patel, K.: Bond rating methods: Comparison and validation. *The Journal of Finance* **30** (1975), 631-640.
[4] De Laurentis, G., Maino, R. and Moletni, L.: *Developing, validating and using internal ratings. Methodologies and case studies*. John Wiley & Sons Inc., Chichester, 2010.
[5] Hosmer, D. W. and Lemeshow, S.: *Applied logistic regression*. John Wiley & Sons Inc., New York, 2000.
[6] Pinches, G. and Mingo, K. A.: A multivariate analysis of industrial bond ratings. *The Journal of Finance* **28** (1973), 1-18.
[7] Rencher, A. C.: *Methods of multivariate analysis*. John Wiley & Sons Inc., New York, 2002.
[8] Tabachnik, B. G.: and Fidell, L. S. *Using multivariate statistics*. Pearson Education, Inc., Boston, 2007.
[9] Waagepetersen, R.: A statistical modeling approach to building an expert credit rating system. *The Journal of Credit Risk* **6** (2010), 81 – 94.
[10] Westgard, S. and Wijst, N.: Default probabilities in a corporate bank portfolio: A logistic model approach. *European Journal of Operational Research* **135** (2001), 338 – 349.

The relationship between monetary and financial stability: Evidence from Central and Eastern European countries

Anca-Elena Nucu[1]

Abstract. The academic literature and recent experience of international crisis converge to the idea that monetary stability is a necessary prerequisite, but not sufficient for ensuring financial stability. The purpose of this paper is to investigate the nexus between monetary stability and financial stability, in the experience of several Central and Eastern European countries (Bulgaria, Czech Republic, Hungary, Latvia, Lithuania, Poland, Romania), during 2004M01-2012M02. Using a Vector Autoregressive model, we analyze the impact of monetary policy interest rate (proxy variable for monetary stability) upon the short term interest rates, exchange rates, share prices and loan to deposit ratio (proxy variables for financial stability). We want to test if key policy rate instrument is conducive to financial stability. The main findings of the paper emphasize that monetary policy interest rate is conducive to financial stability only in the case of Czech Republic and Poland, countries with a high degree of monetary policy autonomy. In the case of Bulgaria and Lithuania, the changes in European Central Bank (ECB) refinancing interest rate are not in accordance with specifically countries conditions. In Latvia, there are similarities between the monetary condition and the currency board strategy and, in Hungary and Romania, the interest rate instrument, used for inflation targeting, is not conducive to financial stability.

Keywords: monetary stability, financial stability, key interest rate, emerging markets, vector autoregressive model.

JEL Classification: C58, D53, E43
AMS Classification: 62M10, 91B82

1 Introduction

The role of central banks in ensuring and maintaining financial stability was reconsidered, after the negative profound implications of the international economic and financial crisis on the real economy. The academic literature converges to the idea that monetary stability is a necessary prerequisite, but not sufficient for ensuring financial stability.

The purpose of our paper is to analyze the nexus between monetary stability and financial stability, in the experience of euro area candidate countries (Bulgaria, Czech Republic, Latvia, Lithuania, Hungary, Poland, Romania), before and during the financial crisis from 2008. Using a Vector Autoregressive model, we analyze the impact of monetary policy interest rate (proxy variable for monetary stability) upon the short term interest rates, exchange rates, share prices and loan to deposit ratio (proxy variables for financial stability).

Our empirical results show that the impact of monetary policy interest rate on financial variables depends on country specific conditions.

The paper is organized as follows. The next section briefly surveys the major contributions of the literature review. Section 3 lays out the data and the methodology used. Section 4 evaluates the empirical results. Section 5 concludes.

2 Literature review

The academic literature dedicates numerous studies to the nexus between monetary policy and financial stability, but there is still no clear consensus whether there are trade-offs or synergies between them. Assenmacher-Wesche and Gerlach [2] and Hunter et al. [8] argue that an explicit and proactive response of the monetary authorities to the financial imbalances is neither desirable nor feasible. On the other hand, Manolescu [13], Albulescu [1], Vinals [14] and Ingves [9] show that the potential costs of financial instability are large enough to justify a proactive approach of monetary policy.

[1] "Alexandru Ioan Cuza" University of Iasi, Department of Economic and Business Administration, Carol I Avenue, no.11, Iasi, Romania, nucu.anca@yahoo.com.

The major challenge comes from the fact that financial stability is a multidimensional concept and there is no widely accepted definition or a standard measurement framework. As opposed to monetary stability which is a clearly defined objective, Albulescu [1] states that financial stability may include, but it is not limited to a numerical objective.

In this paper, financial stability is defined in terms of changes in the structure of interest rates, exchange rate, share prices and bank loan-deposit ratio. Various authors have examined the impact of monetary policy via key interest rate on financial variables. From the empirical standpoint, the impact of monetary policy on key macroeconomic variables is analyzed within Vector Autoregressive (VAR) approach, using constraints and impulse response analysis. For example, Granville and Mallick [6] examine, in the experience of EMU between 1994 and 2008 the response of the term structure of interest rates, share prices, exchange rates, property price inflation and the deposit-loan ratio to changes in the consumer price level and ECB policy rate and they they found that there is a pro-cyclical relationship between monetary and financial stability in the long run. Honda and Kuroki [7] investigate the effects of changes in the policy target variable on stock prices and the term structure of interest rates in Japan and find that changes in the surprise component of the target variable significantly affect both intermediate-term and long-term interest rates. Moreover, they find an inverse relationship between the target rate and stock prices. Also, Ioannidis and Kontonikas [10], Bjorland and Leitemo [3], Li et al. [11] show that an unexpected increase in key interest rate drop share prices down, in developed countries. On the other hand, the impact of stock returns on monetary policy decisions has not received special attention from an empirical standpoint.

Much less, hovewer, has been written about the optimal monetary instrument to maintain financial stability. Goodhart et al. [5] assess the choice between adopting a monetary base or an interest rate setting instrument for prudential purposes. The authors suggest that the interest rate instrument is preferable, since during times of a panic or financial crisis the central bank automatically satisfies the increased demand for money.

3 Data and Methodology

The impact of monetary policy on financial variables is analyzed within the VAR framework, in order to capture the dynamic interactions. We want to test if the key policy rate used for monetary stability is conducive to financial stability in the experience of several Central and Eastern European Countries: Bulgaria, Czech Republic, Hungary, Latvia, Lithuania, Poland and Romania.

3.1 Data

For each country, a five dimensional VAR model is defined with monetary policy interest rate, short term interest rate-3 months, log of exchange rate measured as local currency per EURO, log of share prices and loan to deposit ratio for the banking system. Particularly, for Bulgaria and Lithuania, we took the ECB refinancing interest rate, due to their currency board monetary policy strategy, which implies the lack of controlled monetary instruments. The monetary policy interest rate is the main rate established by the central bank for the primary instrument of monetary policy, around which fluctuates the overnight market interest rate.

We have selected a proxy variable for the main markets which incur risks that may affect the stability of the domestic financial system in selected countries, as follows:

- short-term interest rate for the money market. A prudent monetary policy stance try to bring money market rates to normal levels in order to consolidate the favorable conditions for the sustainable resumption of lending to the private sector and economic rebound. If the short term interest rate appears to be disconnected from the key interest rate, this is a useful insight of a potential imbalance in the financial system.
- exchange rate measured as local currency per EURO for foreign exchange market. The exchange rate stability plays an important role in developing foreign investments and minimizes the currency risk of debtors. The negative comovement between key interest rate and exchange rate is desirable under a monetary policy shock. An exogenous interest rate increase by the central bank which does not lead to an appreciation of the local currency denotes an imbalance in the financial system.
- share prices for capital market. The speculative bubbles of asset prices may degenerate into a financial crisis. From the perspective of a central bank, an unexpected increase in key interest rate should drives share prices down.
- loan to deposit ratio for banking market. A credit boom which it is not accompanied by an increase in the level of deposits (which reflects the confidence in national currency) indicates a potential imbalance in the financial system and the fact that households and companies face the problem of informational asymmetry.

Therefore, a positive monetary policy shock should leads to a decrease in loan to deposit ratio for the banking system.

Monthly time series data ranging from 2004M01 to 2012M02 have been used. We divide the period analyzed in two subperiods as follows: 2004M01- 2008M07 and 2008M08- 2012M02 in order to investigate the impact of monetary policy interest rate on financial variables mentioned above during the financial crisis.

All series are provided by the IMF-IFS statistics (via Datastream) and the precise Datastream mnemonics are available on request.

3.2 Methodology

Let Y_t be a VAR model of order p with the following form:

$$\mathbf{Y_t = \upsilon + A_1 Y_{t-1} + ... + A_p Y_{t-p} + u_t}, \tag{1}$$

where Y_t is a (K*1) vector of endogenous variables, ν is a K*1 vectors of intercepts, A_i are the (K*K) fixed VAR coefficient matrices and $u_t=(u_{1t},...,u_{kt})'$ is an unobservable error term. It is assumed to be a zero-mean independent white noise process with time-invariant, positive definite covariance matrix $E(utut')=\sum u$. K is the number of time series variables and it is equal to 5: monetary policy interest rate, money market interest rate- 3 month, log of exchange rate, log of share prices and loan to deposit ratio for the banking system, computed as follows:

$$\mathbf{Loan\_to\_deposit\_ratio_j = Total\_loans_j \Big/ Total\_deposits_j * 100} \tag{2}$$

where j= Bulgaria, Czech Republic, Hungary, Latvia, Lithuania, Poland and Romania.

The process is stable if

$$\mathbf{det(I_k - A_1 z - ... - A_p z^p) \neq 0, \ for\ |z| \leq 1} \tag{3}$$

which means that the polynomial defined by the determinant of the autoregressive operator has no roots in and on the complex unit circle.

A situation of special interest arises if some of the variables taken into consideration are driven by a common stochastic trend. We have tested the number of cointegrating relationships by using the maximum likelihood methodology of Johansen and we have performed the cointegration test based on all pairs of series, before turning to the five dimensional system, because the cointegration rank test tend to have relatively low power when it is applied to higher dimensional system.

If cointegrating relations are present, the VAR form is not the most convenient model setup and, therefore, we consider specific parameterizations that support the analysis of the cointegration structure. In this case, Vector Error Correction model (VECM) with the following standard representation:

$$\mathbf{\Delta Y_t = \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + ... + \Gamma_{p-1} \Delta Y_{t-p+1} + u_t} \tag{4}$$

is more convenient, where

$$\mathbf{\Pi = -(I_k - A_1 - ... - A_p)} \tag{5}$$

$$\mathbf{\Gamma_i = -(A_{i+1} + ... + A_p), \quad for\ i=1,...,p-1}. \tag{6}$$

The VECM is obtained from the levels VAR form by subtracting Y_{t-1} from both sides and rearranging terms. $\prod Y_{t-1}$ is the only one that includes I(1) variables and must also be I(0). Thus, it contains the cointegrating relation. The $\Gamma_1,..., \Gamma_{p-1}$ are often referred to as short run parameters and $\prod Y_{t-1}$ is sometimes called the long term part.

Against the background that the variables may be cointegrated, we follow the standard practice in this line of the literature and specify VAR models for the levels of the variables. Brüggemann and Balabanova [4] state that this fact avoids the false cancelation of long-run relationship between the variables.

For checking if the selected VAR model or VECM provides the best representation of the time series set, we have tested against the residual autocorrelation, nonnormality, ARCH effects and parameter instability.

4 Empirical results

The unit root analysis, according to Augmented Dickey-Fuller (ADF) and Philips-Perron tests, indicates that the unit root hypothesis cannot be rejected for the all the considered time series and that most of them can be characterized as integrated of order 1, I(1).

Results in our estimation (table 1) confirm a direct relationship between monetary policy interest rate and short term market interest rates in all countries, before and after the financial crisis, except Bulgaria where, during 2004-2008M07, there is an inverse relationship between these two variables. This means that, before the financial crisis, domestic interest rate was disconnected from the ECB interest rate in the short run. Contrary to the results for Bulgaria, in Lithuania, domestic short term interest rate track closely the ECB interest rate, which offer some insights that the velocity of the convergence process, taking into consideration the anchor interest rate value, is greater in Lithuania compare to Bulgaria.

The empirical results show a significant relationship between key interest rate and exchange rate in Czech Republic, Poland, Romania during 2004M01 and 2008M07, meaning that an increase in key policy rate leads to the appreciation of local currencies (CZK, PLN, RON) against the single european currency. Taking into consideration the second subperiod, we find an inverse significant nexus between these two variables only in the case of Poland and Czech Republic, according to traditional macroeconomic theory. Also, in the second subperiod, there we found no significant relationship between monetary policy interest rate and exchange rate in Hungary and Romania. Note that the currency board strategy implies that the exchange rate of BGN and LTL against EURO is constant, therefore, from an empirical standpoint, the analysis has no coherent interpretation. Bulgaria opted for a tight peg against the single currency and Lithuania, by joining Exchange Rate Mechanism II (ERM II), in 2004, retains a stable exchange rate of the litas against the anchor (EUR). Moreover, Latvia adopted a fixed peg on the SDR, and, in terms of exchange rate regime, Minea and Rault [12] show that there are similarities with currency board. As litas, the Latvian lats was included in ERM II. For these countries, the exchange rate enters in the model as exogenous variable.

The empirical results show that the degree of capital market development influences the impact of key interest rate on stock prices. Therefore, we find an inverse significant relationship between monetary policy interest rate and stock market indices, during the whole sample, only in the case of Poland and Czech Republic. Taking into consideration the first period, there we found a direct relationship between the above mentioned variables in Lithuania, which means that ECB interest rate was not an efficient instrument of intervention for mitigating the excessive accumulation of financial imbalances before the financial crisis. In the aftermath of the financial crisis, we remark an improvement of interest rate as instrument of intervention on Latvian and Lithuanian stock market indices. An increase with one unit of key interest rate leads to a decrease of stock prices with 3.013155 and 5.849623 in Latvia and, respectively Lithuania, at a significance level of 5%. For Bulgaria, Romania and Hungary, there we found no significant nexus between these two variables at any degree of freedom, neither before, nor after the financial crisis.

Country	Financial variables	Log of Index 2004M01-2008M07		Log of Index 2008M08-2012M02	
		Coefficient	t-Statistic	Coefficient	t-Statistic
Czech Republic	Interbank rate 3 Month	0.514442	[3.35492]*	0.431099	[-3.18026]*
	Exchange rate CZK/EUR	-1.239145	[-2.37053]*	-1.738963	[3.10867]*
	Prague Stock Exchange Index	-4.885486	[-1.68979]*	-12.70421	[-1.39338]**
	Loan to Deposit Ratio	-80.13602	[-2.69697]*	-27.47611	[-1.47720]**
Bulgaria	Interbank Rate 3 Month	-0.729257	[-1.66371]**	0.773021	[2.90507]*
	Exchange rate BGN/EUR	N/A	N/A	N/A	N/A
	Sofia Stock Exchange index	-9.212778	[-0.22531]	5.289994	[0.64772]
	Loan to Deposit Ratio	-3768.237	[-1.77916]*	584.6438	[2.28306]*
Latvia	Interbank rate 3Month	0.020782	[2.03463]*	0.690576	[1.30260]**
	Exchange rate LVL/EUR	N/A	N/A	N/A	N/A
	Riga Stock Exchange Index	4.415330	[0.99734]	-3.013155	[-2.70268]*
	Loan to Deposit Ratio	-4.913834	[-1.85268]*	2.769160	[1.96383]*
Lithuania	Interbank rate 3 Month	0.424281	[1.98675]*	0.560453	[2.75884]*
	Exchange rate LTL/EUR	N/A	N/A	N/A	N/A

	Vilnius Stock Exchange Index	12.55470	[3.88781]*	-5.849623	[-3.35270]*
	Loan to Deposit Ratio	2.748612	[1.21394]	2.377241	[2.90705]*
	Interbank rate 3 Month	0.382162	[1.57901]**	0.567629	[-2.41943]*
Hungary	Exchange rate HUF/EUR	-4.074477	[-1.24701]	1.984404	[0.56159]
	Budapest Stock index	2.755473	[0.55919]	-2.177802	[-0.47794]
	Loan to Deposit Ratio	9.616931	[0.45135]	-1.701803	[-0.53746]
	Interbank rate 3 Month	0.729278	[4.57557]*	0.480941	[3.47035]*
Poland	Exchange rate PLN/EUR	-0.203040	[-2.96227]*	-0.500823	[-2.82832]*
	Warsaw Stock Exchange Index	-1.842568	[1.31971]**	-7.24402	[2.62834]*
	Loan to Deposit Ratio	-0.156886	[-2.32585]*	-0.241989	[-1.74534]*
	Interbank rate 3 Month	0.243017	[1.56980]**	1.629858	[-2.09022]*
	Exchange rate RON/EUR	-0.004965	[-1.51555]**	-0.000817	[-0.01943]
Romania	Bucharest Exchange Trading Index	0.088034	[0.04350]	-3.695464	[-0.67651]
	Loan to Deposit Ratio	0.408619	[0.75130]	2.954762	[1.80694]*

Notes: *significance level 5%, **significance level 10%

Table 1 Summary of VAR/VECM outcomes

Our empirical analysis regarding the impact of monetary policy on loan to deposit ratio shows a significant inverse relationship between these two variables in Czech Republic, Bulgaria, Latvia and Poland during 2004-2008M07, which means that key interest rate was an efficient instrument of intervention in mitigating the excessive credit growth. Although, the banking system is the core of the financial system, like in the mentioned above countries, for Romania, Hungary and Lithuania, there we found no significant nexus between the variables. These may due to the fact that the presence of foreign capital banks limits the efficiency of monetary policy. Taking into consideration the second subperiod, we find, also, an inverse relationship between key interest rate and loan to deposit ratio in the case of Czech Republic and Poland. Our results suggest that restrictive monetary conditions promote risk-taking among banks in Bulgaria, Latvia, Lithuania and Romania, where an increase in interest rate leads to a raise of loan to deposit ratio. This fact offers some insights that, when interest rate is not controlled by the national central bank, households and companies face the problem of informational asymmetry.

From historical perspective, interest rates were very low during the last years and liquidity trap limited the monetary policy efficiency in selected countries.

The only potential models defect is nonnormality for the residuals of the interest rate equation. The fact that residuals are not normally distributed is not surprising and is not necessarily a signal of inadequate modeling. The main caveat of our analysis is the fact that VAR/VECM models contain many parameters and with short samples, but the start of the estimation sample is governed by data availability.

5 Conclusions

Our paper investigates the nexus between monetary stability and financial stability, in the experience of several Central and Eastern European countries (Bulgaria, Czech Republic, Hungary, Latvia, Lithuania, Poland, Romania), before and after the financial crisis. Using a Vector Autoregressive model, we analyzed the impact of monetary policy interest rate (proxy variable for monetary stability) upon the short term interest rates, exchange rates, share prices, and loan to deposit ratio (proxy variables for financial stability).

Our empirical results suggest the following:

- in Czech Republic and Poland, countries with a high degree of monetary policy autonomy, the key interest rate instrument used for inflation targeting is conducive to financial stability;
- in Bulgaria and Lithuania, countries which lost their monetary policy autonomy, the changes in ECB refinancing interest rate are not in accordance with specifically domestic conditions. We found evidences that before the financial crisis, domestic interest rate in Bulgaria was disconnected from the ECB interest rate in the short run. Our results offers some insights that, when interest rate is not controlled by the national central bank, households and companies face the problem of informational asymmetry. Also, the ECB interest rate is not used accordingly to stock market conditions. Nevertheless, we find some insights that the velocity of the convergence process, taking into consideration the anchor interest rate value, is greater in Lithuania compare

to Bulgaria. Moreover, we remark an improvement of interest rate as instrument of intervention on Lithuanian stock prices, in the aftermath of the financial crisis;

- in Latvia, there are similarities between the monetary condition and the currency board strategy;
- in Hungary and Romania, the interest rate instrument used for inflation targeting is not conducive to financial stability. For Hungary, except the short term interest rate, there we found no significant nexus between key interest rate and the other financial variables, at any degree of freedom, nor before, neither after the financial crisis. In the case of Romania, we found no significant relationship nor between key interest rate and stock prices during the whole period neither between key interest rate and exchange rate after September 2008. Moreover, the restrictive monetary conditions promote risk-taking among Romanian banks. Achieving simultaneous monetary and financial stability calls for refinements of the current monetary and prudential policy frameworks in Hungary and Romania.

Future research is needed to complete our results. A first development is impulse response analysis, under the assumptions that structural shocks are orthogonal. A second development should give interest to the external economic developments in the analysis of the efficiency of monetary policy in selected CEE countries.

Acknowledgements

This work was supported by the the European Social Fund in Romania, under the responsibility of the Managing Authority for the Sectoral Operational Programme for Human Resources Development 2007-2013 [grant POSDRU/107/1.5/S/78342.

References

[1] Albulescu, C.T.: *Stabilitatea sectorului financiar în condiţiile aderării României la U.E.M.* Editura Universităţii de Vest, Timişoara, 2010.
[2] Assenmacher-Wesche, K., and Gerlach, S.: Can Monetary Policy Really Be Used to Stabilize Asset Prices?. In: *The First Global Financial Crisis of the 21st Century*, Center for Economic Policy Research, London, 2008, 163-166.
[3] Bjørnland, H., and Leitemo, K.: Identifying the interdependence between US monetary policy and stock market. *Journal of Monetary Economics* **56** (2009), 275–282.
[4] Brüggemann, R., and Balabanova, Z.: *External Information and Monetary Policy Transmission in New EU Member States: Results from FAVAR Models*, Working Paper, Department of Economics, University of Konstanz, 2012.
[5] Goodhart, C., Sunirand, P., and Tsomocos, D.P.: The optimal monetary instrument for prudential purposes. *Journal of Financial Stability* **7**(2011), 70–77.
[6] Granville, B., and Mallick, S.: Monetary and financial stability in the euro area: pro-cyclicality versus trade-off. *Journal of International Financial Markets, Institutions & Money* **19**(2009), 662-674.
[7] Honda, Y., and Kuroki, Y.: Financial and capital markets' responses to changes in the central bank's target interest rate: the case of Japan. *The Economic Journal* **116** (2006), 812-842.
[8] Hunter, W., Kaufman, G., and Pomerleano. M.: *Asset prices bubbles: The Implications for Monetary, Regulatory and International Policies*, MIT Press,U.S.A, 2005.
[9] Ingves, S.: *Central bank governance and financial stability*, A report by a Study Group, Bank for International Settlements, 2011.
[10] Ioannidis, C., and Kontonikas, A.: The impact of monetary policy on stock prices. *Journal of Policy Modeling* **30** (2008), 33–53.
[11] Li, Y., Işcan, T., and Xu, K.: The impact of monetary policy shocks on stock prices: Evidence from Canada and the United States. *Journal of International Money and Finance* **29** (2010) 876-896.
[12] Minea, A., and Rault, C.: *Monetary policy transmission: old evidence and some new facts from Bulgaria*, Bulgarian National Bank, Discussion Paper, **67**(2008).
[13] Manolescu, G.: *Politica monetară în contextul globalizării*, Editura Universitară, Bucureşti, 2009.
[14] Viñals, J.: *Central Banking Lessons from the Crisis*, International Monetary Fund, 2010.

Construction and application of scoring models

Vladěna Obrová[1]

Abstract. This paper focuses on the definition and practical application of scoring models that are used not only in the bank sector, but also by other institutions to assess whether the client can get a loan or not. If there is an incorrect decision due to wrong definition in this model then the institution is exposed to risk of loss. Therefore when any application for the loan is posted, institution needs to do the best quantification of the risks. Only after that is able to decide, on the basis of available information, about the applicant. There will be used logistic regression method in the process of constructing the models in this paper. The output of this method has to estimate the conditional probability of client repayment of the loan with given characteristics. One of the most studied characters of the scoring model is the ability of diversification represented by the Lorenz curve and quantified by Gini coefficient.

Keywords: scoring model, Gini coefficient, Lorenz curve, credit risk.

JEL Classification: C190, G210
AMS Classification: 62P20

1 History

The history of credit scoring only started some 50 to 70 years ago. In 1936 Ronald Aylmer Fischer for the first time approached the issue of identification of groups within the population. In 1941 David Duran for the first time recognised that Fischer ´s techniques might be used for distinction between good and bad loans. In 1950s Bill Fair and Earl Isaac established the first advisory office. Considerable progress of credit scoring started with the launch of credit cards in 1960s and growth of information technologies. In 1980s a reliable method of logistic regression and linear programming was introduced. Recently expert systems and neuron networks began to be used like in other risk management areas apart from scoring models. There are also other developing methods such as the approximation function, the Bayesian method, the classification trees, the genetic algorithms and others. [2]

2 Scoring model – assumptions

The following article deals with construction of a scoring model. To be able to discuss the issue I need to define the basic terms. The most important term related to scoring models requiring definition is risk. The term risk is connected with 2 more terms: The first is **uncertain result**, which means that there must be at least two variants of the solution for if you know for sure that a loss will be incurred then you cannot speak of a risk. The second is **undesirability** of at least one of the two possible results. Risk is often understood as the danger of occurrence of a certain loss. [9] Theory of finance usually defines as a risk volatility of a financial quantity (portfolio value, earnings etc.) around the expected value as a consequence of changes of numerous parameters.

2.1 Risk classification – credit risk

The notion of scoring model is related not only to risk but also with classification of risks. There are financial and non-financial risks depending on whether the asserted risk factor causes financial loss or not. In the case of the scoring model the risk will always be the financial risk. Financial risk comprises the relation between a subject (individual or organisation) and assets or expected income which may be lost or deteriorated.

Financial risk is usually affected by three factors:

- subject exposed to the risk of loss;
- assets or income whose value reduction, destruction or change of ownership cause the financial loss;
- danger (threat) that may cause the loss. [9]

[1] BUT, Faculty of Business and Management, Kolejní 2906/4, Brno, 612 00, obrova@fbm.vubtr.cz

Financial risks can be divided **to credit risk, operation risk, liquidity risk** and **market risk**. The terms will be defined below, with special focus on credit risks. Operation risk is connected with operation drawbacks or errors causing loss. This risk is defined by the Bank for International Payments with registered offices in Basel as the risk of direct or indirect losses caused by inappropriate or unsuccessful internal processes, employees or systems or external events. [11]

The liquidity risk is related to the ability of the company in question to cater for all its due liabilities at any moment, which in the case of a financial institution means to be able to pay out the due deposits of its clients at any time and in the required form. Market risk follows from changes of market prices and their impact on earnings (equity value) of the company.

The last financial risk is **credit risk,** which ranks among the basic financial risks for loan granting is common and recently represents one of the most frequently discussed themes in all economic areas. The risk is also called loan risk and as one of the current trends is a trend towards multiple loans acquired by clients, the other party (the bank, the company) needs to define certain parameters to assess whether the client will be able to fully refund the loan within the agreed maturity deadline. The following table shows history of loans in the Czech Republic:

SELECTED LOAN TYPES	ABSOLUTE AMOUNT (CZK BILLION) as of 31 March 2012	YEAR-ON-YEAR CHANGE as of 31 March 2012
Total volume of loans	2304.5	5.72%
Loans to non-financial institutions	832.2	5.14%
Short-term (up to 1 year inclusive)	266.2	7.62%
Medium term (1 – 5 years)	150.3	3.60%
Long-term (over 5 years)	415.6	4.16%
Retail loans	1012.5	4.79%
Debit balances of current accounts	12.7	-1.09%
Liabilities from credit cards	24.7	1.81%
Consumer loans	157.4	-2.34%
Housing loans	777.4	6.48%
Other	40.2	6.53%
Loans to traders	37.3	-6.36%

Table 1 History of loans [10]

At present the conditions for loan provision become more and more stringent for the reason of past non-payments by clients. These procedures follow from historic data. Credit risk is one of the most frequent risks and that is why the banks/financial institutions pay a lot of attention to it.

The important indicators checked by the bank before providing a loan to the client include the client ´s creditworthiness, on the basis of which the bank specifies the amount of the loan and the conditions under which the loan may be granted to the client in question. The amount of net income reduced by other loan, overdraft or credit card repayments determines the amount of the mortgage loan to be provided to the applicant. For the financial institution to be able to decide whether to grant a loan to the given client or not the client creditworthiness must be complemented with assessment of other factors as well. Scoring models are used for this purpose of parameter benchmarking and client assessment in practice. The scoring is expressed by a real number, the score. You may also say that the client ´s score means numerical expression of financial reliability of the client. Higher scores are granted by the model to clients with higher creditworthiness while clients with low creditworthiness receive low scores. The higher score - the lower risk of the loan non-repayment. Thanks to imperfections sometimes a good client receives a lower score than a risky client. [3]

2.2 Credit scoring

Scoring model is then a specialised risk rate trying to “squeeze” all risks of an institution/client into a single number. Credit scoring focuses on which clients will not repay the loan and which clients will not pay for provided services. I will continue to deal with just the credit scoring, also used for detection of fraud, on loan application approval/rejection or pre-approval.

Methods of **credit scoring** form a standard part of risk management by financial institutions. On the basis of this model each applicant for a loan receives a score. The score means assessment of the client. The higher the score than better is the client. Sometimes the score is represented by a probability estimate whether the given client is to repay the loan or not. On the basis of the score the institution decides about the conditions under which the loan will be provided. In the case of new clients there is the so called application scoring on the basis of which the bank decides whether to provide a loan to the applying client at all. **Application scoring** models are used by loan institutions to evaluate creditworthiness of potential clients applying for credit product. The aim of scoring models is to classify applicants into two groups: the ones who will not default and the ones who will default. [6] The problem of the application scoring is that the quantity of the assessed clients is different from the quantity of the clients to whom the loan has actually been provided and who become the basis of the application scoring model.

Current clients of the bank are assessed on the basis of the so called behavioural credit scoring allocating a score to all clients, not only to the clients currently applying for a loan or service. As follows from the name of the scoring this model is based on behaviour of the clients of the assessing institution. On the basis of this score selected clients are sent marketing offers of loans or increased overdraft as they have been assessed by the system as clients able to repay higher loans of any kind. In 2007 has Goran Klepac introduced a new methodology of temporal influence measurement (seasonal oscillations, temporal patterns) for behavioural scoring development purposes. His work shows how significant temporal variables can be recognised and then integrated into the behavioural scoring models in order to improve model performance. [5]

Application of the scoring is conditioned by availability of a large volume of homogeneous data, which is one of the reasons why these models have been used by the bigger organizations for unified loans (credit cards, instalments). An important thing is that the scoring model is not to explain the risk, but just to predict it.

Scoring model is usually based on the database of the existing clients to whom a loan has been provided, together with information which clients managed to repay the loan in full. For the sake of simplicity a good client is a client who repaid the loan in full and in time and under the agreed conditions, while a bad client is a client who failed to meet one of its liabilities. Non-repayment is often also called default.

3 Scoring model - construction

The model construction will be based on logistic regression, the difference between logistic regression and the linear regression is, that we did not try to estimate a value of explained variable, but the probability, that the object belong to the one category. [8]

The outputs of the logistic regression method include an estimate of conditioned probability of repayment by the client with the given characteristics. One of the most important analysed features of scoring model is the ability to diversify, i.e. the good/bad client resolution. This feature is expressed in practice by Lorenz curve and quantified by Gini coefficient.

The model construction will be based on logistic regression. The outputs of the logistic regression method include an estimate of conditioned probability of repayment by the client with the given characteristics. One of the most important analysed features of scoring model is the ability to diversify, i.e. the good/bad client resolution. This feature is expressed in practice by Lorenz curve and quantified by Gini coefficient.

Where the scoring model construction is based on the logistic regression model, the simplest way is to start from the binary explained variable Y, which is depends on the explaining variables x_i'. The Y variable, with probability π, acquires value 1, and with probability $(1 - \pi)$, value 0. There is also the explaining variable vector [4]

$$x_i^{'} = [x_{i1}, x_{i2} \dots, x_{ik}\ ], i = 1{,}2 \dots n \tag{1}$$

representing *i-th* combination of the explaining variables $\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_\mathrm{k}$, $X_1, X_2, \dots, X_k$, and then $i - th$ conditioned classification of the Y quantity is alternative [10] with parameter written as $Y \sim Alt(\pi)$ (and median value of the Y quantity, $\mathrm{E}(\mathrm{Y_i}) = \pi_\mathrm{i})$ π_i a pravděpodobnostní funkcí,

$$P(y_i|\pi_i) = {\pi_i}^{y_i}(1 - \pi_i)^{1-y_i} \tag{2}$$

When selecting variables for vector x this probability function is used if for different vectors of x_i values of quantities $X_1, X_2, \dots, X_k$ the conditioned probability ratio of quantity Y (given by parameter π_i) s is the same. Then the Y quantity does not depend on these variables. But if different combinations of values result in different probability ratios π_i, then a certain type of relation between Y and these explaining variables can be assumed and their illustration by the regression model.

3.1 Explaining variables

The explaining variables are expressed with the help of the "score table". The explaining variables may be both quantitative and qualitative. Quantitative variables may be expressed in numbers and are subdivided to discrete (number of client ´s children, number of persons living in the same household as the client) and continuous (monthly income, monthly expenditures). However, most variables are qualitative and cannot be expressed in numbers, such as the highest achieved education, marital status, refunding of previous loans. The score table may include other variables in addition to ownership status, age, loan purpose, and litigation costs. Most variables included in the score table are clearly linked to the risk of non-repayment of the loan. Some variables give an ideal about the client ´s stability. These for example include the length of residence on the current address, or the time with the current employer. Other variables define financial sophistication of the client, such as whether the client has a checking account or a savings account, if h owns any credit cards, and how long the client has been with the current bank. Other variables show the consumer ´s resources. These include ownership status, employment, number of children etc. [3]

These variables are divided into groups by their meaning (such as gender, age, income etc.). Each group consists of several categories. It is further assumed that each client belongs to a single category within each group. To be able to introduce these variables in the model one needs to allocate dummy (binary) variables to them acquiring the value of 0 if the subject does not belong to the category or 1 if the subject falls within the category. Then each binary variable x_j^i of the x set indicates whether the client belongs to the respective $j-th$ category of $i-th$ group ($x_j^i = 1$) či nikoliv ($x_j^i = 0$).

3.2 Logit

There is looked for the relation between π_i and the values of vector x on the basis of the assumption that Y is a binary variable acquiring values 0 or 1 only, and π_i is the probability value from the interval $[0, 1]$. If linear regression is used then the variable can generally acquire any real value. That is why the odds function (or chance) is defined as

$$\text{odds} = \frac{P(Y=1)}{P(Y=0)} = \frac{\pi}{1-\pi}. \tag{3}$$

Thus defined function already acquires a value from the interval $[0,\infty]$. To be able to obtain a function acquiring a value from the whole definition range R it is necessary to calculate the logarithm of the function. Then the function is defined as a logit.

$$\text{logit} = \ln(\text{odds}) = \ln\left(\frac{\pi}{1-\pi}\right) \tag{4}$$

If $\text{logit} = \beta x'$, where $x' = (1, x_1, x_2, .. x_k)\ \beta = (\beta_0, \beta_1, \dots, \beta_k)$, then the logistic regression relation results:

$$\pi = \frac{e^{\beta x'}}{1 + e^{\beta x'}}. \tag{5}$$

The parameters in the linear combination of the explaining variables express the transformed median value of the explained variable – the logit. Logit is the logarithm of the quotient $\pi/1-\pi$ expressing the chances (odds) for the Y quantity to acquire the value of 1. The parameter β_0 expresses the size of the logit for zero values (categories) of all explaining variables. For $\beta_0 = 0$ the chances that $Y = 1$ are one to one, or $\pi = 0.5$. Positive values of the parameter β_0 mean that these odds are higher than one ($\pi > 0.5$) and negative values mean that they are lower than ($\pi < 0.5$). The logit may change in relation to one or more variables. The rate of the change is expressed by parameters $\beta_j, j = 1, \dots k$. In the case of a unit change of a *j-th* explaining variable (on condition that the other variables remain unchanged) the odds are that $Y = 1$, $e^{\beta j}$ times as big. P The maximum plausibility method is applied to logistic regression parameter estimates. [7]

3.3 Odds variables

There is also the $odds$ variable (chances), expressing the ratio of the good clients (G) to the bad clients (B) across the database.

$$odds = \frac{|G|}{|B|}. \tag{6}$$

For the individual features *j* of the individual groups *i* the variables $odds_j^i$, the chances of the feature, are defined as the ratios of the relevant numbers of good and bad clients in the individual categories.

$$odds_j^i = \frac{|G_j^i|}{|B_j^i|} \tag{7}$$

And finally there is the variable $odds\ ratio$, identified as OR_j^i. This variable expresses relative chances of the client within the given category to repay the loan. A value below 1 means that the client ´s chance to repay the loan is under-average, while high values on the other hand show above-average chances.

$$OR_j^i = \frac{odds_j^i}{odds} \tag{8}$$

The main purpose of the credit risk models is to estimate for each potential client with characteristic x o the value of the theoretical characteristic $odds(x)$. In practice it is not recommended to estimate the function $odds(x)$ as the ratio of good to bad clients with characteristic x. on condition of independence of the client in the data set this value will be estimated with the help of the following equation:

$$odds(x) = odds \prod_{(i,j)\in Z} (OR_j^i) x_j^i. \tag{9}$$

That is as the product of the total $odds$ and the respective OR_j^i of the categories where the potential client belongs. Allocations of various weights to the factors in the previous equation can result in different general models.

4 Diversification ability

The ability to diversify, i.e. to separate good clients for bad clients, is one of the most important analysed values of the scoring model. In the ideal case we would like to find a model in which the scoring boundary s_0, would clearly separate all bad clients by allocation of a score lower than s_0 and good clients by allocation of a score higher than s_0. In such a model we would be able on the basis of the calculated score to relatively reliably assess whether the client appears good or not. In practice, however, there is no scoring function faultlessly grasping quality of all clients in the database. There will certainly be clients with a low score who still manage to repay their loans and on the other hand clients who did not repay despite their high scores. The scoring function then only approximately divides the clients to good and bad ones. Quality of a scoring model with regard to its ability to diversify in then assessed according to how well the score is able to separate good clients from bad clients.

4.1 Gini coefficient and Lorenz curve

For graphic representation of the ability to diversify there is for example the Lorenz curve and for numerical quantification the Gini coefficient. Lorenz curve is used not only for graphic representations of scoring models. It also demonstrates the ability of the models to distinguish between good and bad clients. The curve is based on definition of distribution functions of good and bad clients. Gini coefficient is a numerical characteristic of the diversification ability of a scoring model. In the case of credit scoring the Gini coefficient is a benchmark showing how well the score card is able to distinguish between good and bad clients. The final result is the value representing the area under the Lorenz curve, illustrated on the Figure 1.

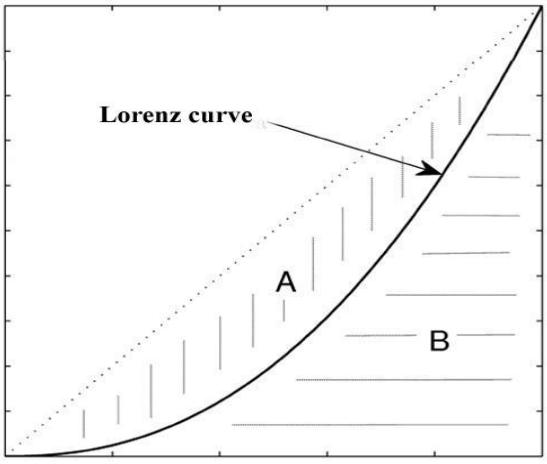


Figure 1 Lorenz curve and Gini coefficient

Let us define for each score value “s” the distribution function of the good client scores FG as the probability that the good client will have the score lower than “s” and the distribution function of the bad clients scores FB at the probability that the bad client will have the score lower than "s". The higher the diversification ability of the model is illustrated by the closer curve to the edges of the given square.

Gini coefficient is defined as the ratio of the oriented area between the Lorenz curve and the diagonal of the unit square (A) to the total area (A+B), i.e.

$$GC = \frac{A}{A + B} \tag{10}$$

Gini coefficient can acquire values from -1 to 1, with values close to 1 representing ideal diversification and values close to 0 representing zero diversification ability and negative values representing opposite classifications of the scoring function (the curve is sagging upwards). Therefore we look for scoring functions with the Gini coefficient as high as possible.

5 Conclusion

The purpose of this article was to describe theoretical assumptions for the construction of scoring models. The essay further draws links between the scoring model and risk, both credit risk and other risk categories. Construction of these models is described in detail including their diversification ability.

Thus defined models can also be used in other sectors apart from the banking sector and can therefore be applied to other areas related to the life of the institution as long as the institution possesses data necessary for the model construction.

Scoring models are irreplaceable for decisions when to lend money to a client and when not, but cannot remain the sole criterion for this decision, as the client can be classified incorrectly, i.e. a client able to repay the loan might be classified in the risky client category and vice versa. And the construction and use of the models should not replace the human factor altogether. At present there are numerous software systems able to serve to other than just banking institutions in using these models in practice.

References

[1] Anděl, J: *The basis of mathematical statistics*. Matfyzpress, Praha, 2007.
[2] Anderson, R: *The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation.* Oxford University Press, New York, 2007, p. 731.
[3] Edelman ,D., Thomas, L. and Crook, J.: *Credit scoring and its applications*. Society for Industrial and Applied Mathematics, Philadelphia, 2002, p. 248.
[4] Hosmer, D. W. and Lemeshow, V.: *Applied Logistic Regression*, New York: John Wiley & Sons, 2000.
[5] Klepac, G.: Integrating Seasonal Oscillations into Basel II Behavioral Scoring Models. *Financial Theory and Practice*, vol. 31, no.3, Zagreb, 2007, pp. 281-291.
[6] Majer, I.: Application scoring: logit model approach and the divergence method compared. *Working Papers 17*, 2006.
[7] Pecáková, I.: Logistic regression with multicategorial explaining variable. *Acta Oeconomica Pragensia*, vol. 15, no. 1, Praha, 2007, pp. 86-96.
[8] Řeháková, B.: Do not be afraid for logistic regression. *Sociologic journal*,, Vol. 36 (No., vol. Vol. 36, no. No.4, pp. 475-492, 2000.
[9] Smejkal, V. and Rais, K.: *Risk management in enterprises and other organizations*. Praha: Grada, 2009, p. 360.
[10] The banks and facts - May 2012, "*Czech bank association*," [Online]. Published 11 05 2012, Available: http://www.czech-ba.cz/data/articles/down_43877.pdf. [Accessed 19 05 2012]
[11] "*EUR-Lex Access to law of European union*," [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52005AE0244:CS:NOT. [Accessed 19 05 2012].

A New(?) k-Shortest Path Algorithm

Stanislav Palúch[1]

Abstract. k-shortests path problem is to find k shortest point-to-point paths in a graph or a digraph. Solution of this problem is useful in many practical cases when we are looking for a sufficiently short path which complies to some additional conditions that cannot be simply formulated. Such problems can be solved by choosing a feasible path among k shortest paths. This paper presents a new simple algorithm for k-shortest paths problem.

Keywords: digraph, shortest path, k-shortest paths problem

JEL classification: C44
AMS classification: 90C15

1 Introduction

Many problems from practical life lead to the problem to find in a network relatively short point to point path which has to fulfill certain special constraints that cannot be simply formulated by means of graph theory. Such problems are very often NP–hard and can be approximately solved by successive enumeration of k shortest paths and consequently by choosing that one from them which complies with given constraints.

The k – shortest path problem is frequently studied in literature. Plesnik presents in [4] a procedure based on prohibiting edges of till now best paths, Yen gives in [5], [6] a deviation algorithm with running time $O(kn(m + n \log n))$ (where n is number of vertices and m is number of edges of undrelying graph) – this worst case assesment has been the best known for long time perion. Algorithm by Gotthilfa, and Lewenstein [1] (issued in March, 2009) runs in time $O(kn(m + n \log \log n))$.

2 Terminology

Since graph terminology is non uniform I present here several essential notions.

A **digraph** (a directed graph) is an ordered pair $G = (V, A)$, where V is a nonempty finite set and A is a set of ordered pairs of the type (u, v) such that $u \in V$, $v \in V$ and $u \neq v$. The elements of V are called **vertices** and the elements of A are called **arcs** of the digraph G.

A $(v1, vk)$ - *walk* in digraph $G = (V, A)$ is an alternating sequence of vertices and arcs of the form

$$\mu(v_1, v_k) = (v_1, (v_1, v_2), v_2, \ldots, v_{k-1}, (v_{k-1}, v_k), v_k).$$

A trivial walk is a walk containing only one vertex. A (v_1, v_k)-**trail** in G is a (v_1, v_k)-walk with no repeated arcs. A (v_1, v_k)-**path** in G is a (v_1, v_k)-walk with no repeated vertices.

An **arc weighted digraph** $G = (V, A, c)$ is an ordered triple where $G = (V, A)$ is a digraph and $c : A \to R$ is a real function defined on the arc set A, the value $c(a)$ for $a \in A$ is called the **weight** of the arc a (or sometimes the **arc-weight**, the **length** or the **cost** of the arc a).

In this paper we will assume that $c(a) > 0$. This condition is fulfilled in many practical applications. The length of the walk $\mu(u, v)$ in a digraph $G = (V, A, c)$ is the total sum of arc-weights of it's arcs, whereas the arc weight is added to the total sum so many times how many times it appears in the walk $\mu(u, v)$.

[1]University of Žilina /Faculty of Management Science and Informatics, Department of Mathematical Methods, Univerzitná 8215/1, 010 26 Žilina, Slovakia, e-mail: stanislav.paluch@fri.uniza.sk

Suppose that $\mu(v_1, v_k) = (v_1, (v_1, v_2), \ldots, (v_{k-1}, v_k), v_k$ and $\nu(u_1, u_l) = (u_1, (u_1, u_2), \ldots, (u_{l-1}, u_l), u_l)$ are two walks and let $v_k = u_1$. The **concatenation** $\mu(v_1, v_k) \oplus \nu(u_1, u_l)$ of walks $\mu(v_1, v_k)$ and $\nu(u_1, u_l)$ is the walk

$$\rho(v_1, u_l) = (v_1, (v_1, v_2), v_2, \ldots, (v_{k-1}, v_k), v_k \equiv u_1, (u_1, u_2), \ldots, (u_{l-1}, u_l), u_l)$$

Denote $V^+(v) = \{w|(v, w) \in A\}$, $A^+(v) = \{(v, w)|(v, w) \in A\}$.

3 Properties of set of shortest k-paths

Let $G = (V, A, c)$ be an arc weighted digraph. The existence of two different (u, v)-paths with the same length could result in difficulties when considering the k-shortest paths problem. If there are two different (u, v)-paths having the same length, we need another priority rule how to determine which one of them is k-th and which $(k + 1)$-th shortest path. This decision can be made in such a way that in the case of two (u, v)-paths with equal length, we can compare them in lexicographicaly order.

Another way how to awoid just mentioned problem is to enumerate the arcs of the set $A - a_1, a_2, \ldots, a_m$ and introduce new arc cost

$$\overline{c}(a_i) = c(a_i) + \frac{\varepsilon}{2^i} \tag{1}$$

For simplicity we will assume whithout loss of generality that there are no two different (u, v)-paths with the same length.

Proposition 1. *Let*

$$\mu_k(u, v) = (u \equiv v_1, (v_1, v_2), v_2, \ldots, v_{r-1}, (v_{r-1}, v_r), v_r \equiv v) \tag{2}$$

be the k-th shortest (u, v)-path. Let $q < r$ and let $\mu(u, v_q) = (u \equiv v_1, (v_1, v_2), v_2, \ldots, v_{q-1}, (v_{q-1}, v_q), v_q)$. Then $\mu(u, v_q)$ is the l-th shortest (u, v_q)-path for some $l \leq k$.

Proof.
Let $\mu_k(u, v)$ be the k-the shortest (u, v)-path, let $q < r$. Then $\mu_k(u, v)$ can be written as concatenation

$$\mu_k(u, v) = \underbrace{(u \equiv v_1, v_2, \ldots, v_q)}_{\mu_l(u, v_q)} \oplus \underbrace{(v_q, v_{q+1}, \ldots, v_r \equiv v)}_{\mu(v_q, v)} = \mu_l(u, v_q) \oplus \mu(v_q, v) \tag{3}$$

If $l > k$, i. e. if $l - 1 \geq k$ then $\mu_1(u, v_q), \mu_2(u, v_q), \ldots, \mu_{l-1}(u, v_q)$ are $l - 1$ shortest (v, v_q) paths. Therefore we have at least $l - 1 \geq k$ paths – namely

$$\mu_1(v, v_q) \oplus \mu(v_q, v), \mu_2(v, v_q) \oplus \mu(v_q, v), \ldots, \mu_{l-1}(v, v_q) \oplus \mu(v_q, v)$$

shorter than $\mu_k(u, v)$ what is in contradiction with the fact that $\mu_k(u, v)$ is the k-th shortest (u, v)-path. □

Colorary. Let $\mu_k(s, w)$ be the k-th shortest (s, w) path having the length equal to d. Then $\mu_k(s, w)$ is concatenation of the following type

$$\mu_k(s, w) = \mu_l(s, v) \oplus (v, (v, w), w), \tag{4}$$

where $\mu_l(s, v)$ is the l-th shortest (s, v)-path, $l \leq k$, $(v, w) \in A$ and $d(\mu_k(s, w)) = d(\mu_l(s, v)) + c(v, w)$.

Definition 1. Let $\mu_k(s, w)$, $\mu_k(s, w)$ be two paths fulfilling (4). We will say that $\mu_k(s, w)$ is **extension** of $\mu_l(s, v)$ and $\mu_l(s, v)$ is **parent path** of $\mu_k(s, w)$.

Definition 2. Let $\mu(s, w)$ be the k-th shortest (s, w)-path in digraph $G = (V, H, c)$. The number k is called the **rank** of the path $\mu(s, w)$. The rank of $\mu(s, w)$ will be denoted as $\text{rank}(\mu(s, w))$.

Denote by $\mathcal{P}$ the set of paths in $G = (V, A, c)$ containing for every $w \in V$ all shortest (s, w)-paths with rank less or equal to K. Every $\mu_k(s, w)$ is uniquely defined by it's last vertex w, last but one vertex v and the l-th shortest (s, v)-path $\mu_l(s, v)$.

Let $\mathcal{D} \subset \mathcal{P}$, let dmax($\mathcal{D}$) be the length of longest path in $\mathcal{D}$, i. e. dmax($\mathcal{D}$) = max$\{d(path)|path \in \mathcal{D}\}$, let $\mathcal{D}$ contain all paths from $\mathcal{P}$ having the length less or equal to dmax($\mathcal{D}$). Denote by $\mathcal{E}$ the set of all extensions of paths from $\mathcal{D}$ with length greater than dmax($\mathcal{D}$).

If $\mathcal{D}$ does not contain all required shortest paths we can add to it next paths by repeating the following procedure. Choose the shortest path from $\mathcal{E}$ denoted by $\mu(s,w)$ and remove it from $\mathcal{E}$. If no (s,w) path is in $\mathcal{D}$, $\mu(s,w) = \mu_1(s,w)$ is the shortest (s,v)-path. If $\mathcal{D}$ contains k-shortest (s,w)-paths where $k < K$, $\mu(s,w) = \mu_{k+1}(s,w)$ is the $(k+1)$-th (s,w) path. In both mentioned cases insert the path $\mu(s,w)$ into the set $\mathcal{D}$ and insert all extensions of the path $\mu(s,w)$ into the set $\mathcal{E}$. If $\mathcal{D}$ contains K shortest (s,w)-paths, throw away the path $\mu(s,w)$.

The question arises whether $\mu(s,w)$ is realy the $(k+1)$-th shortest (s,w)-path. $\mu(s,w)$ is an extension of a path $\mu_l(s,v) \in \mathcal{D}$:

$$\mu(s,w) = \mu_l(s,v) \oplus (v,(v,w),w) \tag{5}$$

Let $\nu(s,w)$ be a (s,w)-path, $\nu(s,w) \notin \mathcal{D}$ and let $d(\nu(s,w)) < d(\mu(s,w))$. Since $\nu(s,w) \notin \mathcal{D}$, dmax($\mathcal{D}$) $< d(\nu(s,w))$. The path $\nu(s,w)$ can be written as

$$\nu(s,w) = (\underbrace{s \equiv w_1,(w_1,w_2),w_2,\ldots,(w_{t-1},w_t),w_t}_{\nu(s,w_t)},\underbrace{(w_t,w_{t+1}),\ldots,w)}_{\nu(w_t,s)} = \nu(s,w_t) \oplus \nu(w_t,s) \tag{6}$$

where t is largest index such that path $\nu(s,w_t) \in \mathcal{D}$.
Therefore the extension $\nu(s,w_t) \oplus (w_t,(w_t,w_{t+1}),w_{t+1}) \in \mathcal{E}$. But the path $\mu(s,w)$ was chosen as the path with the least length in $\mathcal{E}$ and hence $d(\mu(s,w)) < d[\nu(s,w_t) \oplus (w_t,(w_t,w_{t+1}),w_{t+1})] < d(\nu(s,w))$.

Remember that after just described changes the sets $\mathcal{D}$ and $\mathcal{E}$ still keep their properties – namely $\mathcal{D}$ contains all paths from $\mathcal{P}$ with length less or equal than dmax($\mathcal{D}$) and $\mathcal{E}$ contains all extensions of paths from $\mathcal{D}$ longer than dmax($\mathcal{D}$).

This procedure starts with $\mathcal{D} = \{\mu_1(s,s) = (s)\}$ – $\mathcal{D}$ contains only trivial (s,s)-path and $\mathcal{E} = \{(s,(s,w),w)|(s,w) \in H^+(s)\}$.

4 Algorithm

Assume that $V = \{1,2,...,n\}$. Denote

- n – the number of nodes of the digraph $G = (V,A,c)$
- m – the number of arcs of the digraph $G = (V,A,c)$
- s,f – starting and finishing vertex
- K – the number of searched shortest (s,f) paths
- $c(u,v)$ – the length of the arc $(u,v) \in A$

For every vertex $w \in V$ and for $k = 1,2,...,K$ we will compute step by step at most K definitive tripple labels

$$length[w][k], \quad l\_b\_one[w][k] \quad and \quad rankpp[w][k]. \tag{7}$$

having the following meaning:

- $length[w][k]$ – the length of the k-th shortest (s,w)-path $\mu_k(s,w)$
- $l\_b\_one[w][k]$ – the last but one vertex of the k-th shortest (s,w)-path $\mu_k(s,w)$
- $rankpp[w][k]$ – the rank of the parent path of $\mu_k(s,w)$
- $n\_deflab[w]$ – the number of determined definitive labels of the vertex $w \in V$.

The fact that the labels $length[w][k]$, $l\_b\_one[w][k] = 0$ and $rankpp[w][k]$ are defined for some $w \in V$ and k, $1 \leq k \leq K$ implies that k-th shortest (s, w)-path was discovered.

The sequence of computing of labels (7) will start with setting $n\_deflab[s] = 1$, $length[s][1] = 0$, $l\_b\_one[s][1] = 0$ and $rankpp[s][1] = 0$ for the starting vertex s. This means that one shortest (s, s)-path $\mu_1(s, s)$ was discovered with the length 0 and with no parent path. Subsequently, the next the shortest (s, w)-path $\mu(s, w)$ with property

$$\mu(s, w) \notin \mathcal{D} \quad \text{and} \quad \text{rank}(\mu(s, w) \leq K$$

will be inserted into the set $\mathcal{D}$.

This procedure ensures that if labels (7) are defined for some w and k, i.e. if $\mu_k(s, w) \in \mathcal{D}$, then the parent path $\mu(s, v)$ of $\mu_k(s, w)$ is an element of $\mathcal{D}$, too. The labels of $\mu(s, v)$ define parent path of $\mu(s, v)$ etc, etc. The reverse sequence of all vertices of the path $\mu_k(s, w)$ can be determined step by step in this way.

Just mentioned labels fully characterize the set $\mathcal{D}$. The set $\mathcal{D}$ contains all paths with length less or equal than $\text{dmax}(\mathcal{D})$ and with rank less or equal to K.

The set $\mathcal{E}$ was defined as the set of all extensions of paths from $\mathcal{D}$. Every such extension $\mu(s, w)$ can be fully specified by quadruple (w, t, x, k) where w is the final vertex of $\mu(s, w)$, t is the length of $\mu(s, w)$, x is the final vertex of parent paths of $\mu(s, w)$ and k is the rank of parent path $\mu(s, v)$, i.e. $\mu(s, v) = \mu_k(s, v)$ and $\mu(s, w) = \mu_k(s, v) \oplus (v, (v, w), w)$. The set $\mathcal{E}$ will initially be

$$\mathcal{E} = \{(w, c(s, w), s, 1) | w \in V^+(s)\}.$$

Step 0. Initialization.

Set $n\_deflab[s] = 1$, $length[s][1] = 0$, $l\_b\_one[s][1] = 0$ and $rankpp[s][1] = 0$ for the starting vertex s.

For all $w \in V$ such that $w \neq s$ set: $n\_deflab[w] = 0$ (no definitive label is assigned to w).

For all $w \in V$ such that $w \neq s$:
consider all labels $length[w][j]$, $l\_b\_one[w][j]$ and $rankpp[w][j]$ to be undefined

Step 1. Extracting shortest walk from $\mathcal{E}$.

Extract from $\mathcal{E}$ a quadruple (w^*, t^*, x^*, k^*) having the minimum value of the second item t^* in $\mathcal{E}$.

If $n\_deflab[w^*] \geq K$, we need not another shortest (s, w^*)-path. GoTo Step4.

Quadruple (w^*, t^*, x^*, k^*) determines a (s, w^*)-walk $\mu_{k^*}(s, x^*) \oplus (x^*, (x^*, w^*), w^*)$ with the length t^* and last but one vertex x^*.

Examine whether the path $\mu_{k^*}(s, x^*)$ contains the vertex w.
If yes, the walk $\mu_{k^*}(s, x^*) \oplus (x^*, (x^*, w), w^*)$ contains a cycle and therefore it is not a path.
GoTo Step4.

Step 2. Inserting path (s, w^*) into the set $\mathcal{D}$.

Set:

$$\begin{aligned} n\_deflab[w^*] &= n\_deflab[w^*] + 1 \\ k &= n\_deflab[w^*] \\ length[w^*][k] &= t^* \\ l\_b\_one[w^*][k] &= x^* \\ rankpp[w^*][k] &= k^* \end{aligned}$$

Step 3. Inserting extensions of the path (s, w^*) into the set $\mathcal{E}$.

For all vertices $u \in V^+(w^*)$ such that $n\_deflab[u] < K$
insert the quadruple $(u, t^* + c(w^*, u), w^*, k)$ into the set $\mathcal{E}$.

Step 4. Test of ending conditions.
If $n\_deflab[w] = K$ for all $w \in V$, STOP.
If $\mathcal{E} = \emptyset$, STOP.
Otherwise GoTo Step 1.

The k-th shortest (s, f)-path does exist if $n\_deflab[f] \geq k$ after the algorithm ends. The k-th shortest (s, f)-path $\mu_k(s, f) = (s \equiv v, (v1, v2), v2, \ldots, v_{l-1}, (v_{l-1}, v_l), v_l \equiv f$ can be calculated by making use of labels (7) as follows:

$$\begin{aligned}
v_l &= f \\
k_l &= k \\
v_{l-1} &= l_b_one[v_l][k_l] \\
k_{l-1} &= rankpp[v_l][k_l] \\
v_{l-2} &= l_b_one[v_{l-1}][k_{l-1}] \\
k_{l-2} &= rankpp[v_{l-1}][k_{l-1}] \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
v_2 &= l_b_one[v_3][k_3] \\
k_2 &= rankpp[v_3][k_3] \\
v_1 &= l_b_one[v_2][k_2] \\
k_1 &= rankpp[v_2][k_2]
\end{aligned}$$

The same procedure can be used in step 3. a) for checking whether the path (s,wmin) k contains the wertex w.

5 The complexity of proposed algorithm

Recall that $n = |V|$, $m = |A|$, suppose $m > n$. A quadruple (w, t, x, k) can be inserted into $\mathcal{E}$ at most $K.m$ times. In the case that is $\mathcal{E}$ organized as a Fibonacci heap all insertions into the set $\mathcal{E}$ require $O(K.m)$ steps. Before every insertion a cycle occurrence check is necessary. A single cycle occurrence check requires $O(n)$ steps; all cycle occurrence checks will require at most $O(K.m.n)$ steps. Every quadruple can be extracted from $\mathcal{E}$ at most once, a single extraction requires (in the case of Fibonacci heap) $O(\log(K.m))$ steps, all extractions will need at most $O(Km.\log(Km))$ steps. So the complexity of proposed algorithm is $O(Km(\log(Km) + n))$. Let us remark that we get the same complexity if $\mathcal{E}$ is organized as a binary heap.

6 Modification of algorithm for directed acyclic graphs - DAGs

There are many applications where the studied digraph $G = (V, A, c)$ is a directed acyclic graph – DAG. For example – the underlying digraph for CPM and PERT methods is DAG, the digraph used for optimum bus and/or train connections search is also DAG. Since there are no cycles in DAGs cycle check in Step 1. is not necessary – it can be skipped – and therefore the complexity of proposed algorithm is $O(Km(\log(Km))$.

Let us remark that the omitting the Step 1. in propsed algorithm leads in a digraph which is not acyclic to an algorithm which computes K shortest walks.

7 Conclusion

The proposed algorithm was used for bus and/or train connection search problem with great success. This algorithm was used also to compute shortest feasible trail with respect to prohibited maneuvers as presented by Tomáš Majer in his article on conference MME 2010 in České Budějovice in [3].

Acknowledgements

This research was supported by grant VEGA 1/0374/11 – Modelling and optimization of mobility and infrastructure in logistic networks.

References

[1] Gotthilfa, Z., Lewenstein, M.: Improved algorithms for the k simple shortest paths and the replacement paths problems, *Information Processing Letters*, Vol. **109**, Issue 7, (16 March 2009), 352–355.

[2] Martins, E., Q., W., Pascoal, M., M., B.. Santos, J., L., E.: A New Implementation of Yen's Ranking Looples Path Algorithm, *Investigacao Operacional*, (2000)

[3] Majer, T.: Shortest trail problem with respect to prohibited maneuvers *Mathematical methods in economics 2010 : proceedings of the 28th international conference, September 8-10. 2010, part II* České Budějovice, University of South Bohemia, (2010.) – ISBN 978-80-7394-218-2. 418–422.

[4] Plesnik, J.: *Grafov algoritmy – Graph Algorithms*, VEDA Bratislava, (1983)

[5] Yen, J., Y.: Finding k Shortest Loopless Paths in a Network, *Managements Science*, **17**, (1971), 712–760.

[6] Yen, J., Y.: Shortest Paths Network Problems, *Mathematical Systems in Economics*, Heft **18**, Meisennheim am Glan, (1975)

Permanent Income and Consumption

Václava Pánková[1]

Abstract. A theory of consumer spending which states that people will spend money at a level consistent with their expected long term average income is briefly recapitulated. The level of expected long term income then becomes thought of as the level of "permanent" income that can be safely spent. The permanency can be seen as following an adaptive expectation process what enables to compute values of the variable which in fact is unobservable. A marginal propensity to consume is then easy to find by applying OLS to a simple model. Nevertheless, a question arises whether in reality the response of consumption to income is consistent with such a hypothesis. A validation of the permanent income hypothesis is performed using an instrumental variables approach.

Keywords: permanent income, consumption, instrumental variables, parameter restrictions .

JEL Classification: C12, C13, D11
AMS Classification: 62M10, 62N02

1 Introduction

Long – run economic characteristics can be studied from different points of view. As an economic complex, the contemporaneous situation of Czech Republic is described e.g. in [7]. As for the details, there are numerous possibilities. Permanent income and consumption represent a concept which brings a long – run information because of its nature. Introduced by M.Friedman [4] this phenomenon was studied by using different complementary theories. Adaptive expectation is applied e.g. in [2]. Based on rational expectation hypotheses it is elaborated by Hall [5] and Sargent [8], both approaches harmonized by Flavin [3]. Newly, the concept of rational inattention was added and changing the quality of the topic e.g. by Sims [9].

2 Permanency as a concept

The permanency can be seen as following an adaptive expectation process, details e.g. in [2].

A variable Y_t is supposed to split in two unobservable parts: a permanent one and a temporary one

$$Y_t = Y_t^P + Y_t^T .$$

The permanent value is anticipated to subject an adaptive expectation process as

$$\Delta Y_t^P = Y_t^P - Y_{t-1}^P = \lambda(Y_t - Y_{t-1}^P) \quad \text{with} \quad 0 \le \lambda \le 1$$

It means

$$Y_t^P = \lambda Y_t + (1-\lambda) Y_{t-1}^P \tag{1}$$

with the following interpretation. In year *t* a permanent value is a weighted average of an actual one and a previous permanent value. The previous permanent value follows the same schema, so

$$Y_{t-1}^P = \lambda Y_{t-1} + (1-\lambda) Y_{t-2}^P \qquad \text{a.s.o} \tag{2}$$

[1] Department of Econometrics, University of Economics, Prague, Czech Republic, pankova@vse.cz

By a substitution we then have

$$Y_t^P = \lambda Y_t + \lambda(1-\lambda) Y_{t-1} + \lambda(1-\lambda)^2 Y_{t-2} + \lambda(1-\lambda)^3 Y_{t-3} + \ldots \tag{3}$$

what means that a current value has the greatest weight and the weights decline steadily by going back in the past. Constructing (3) under different choice of λ between zero and one we get $Y_t^{P'}$s in different variants.

Being interested in a final relation, say

$$C_t = \beta_0 + \beta_1 Y_t^P + u_t \tag{4}$$

we can estimate the model according to the variant Y_t^P. We than choose such a λ and relating Y_t^P which produces a best fit of (4) according to the *R*-squared.

3 Permanent income and consumption

Following an idea of Friedman (e.g. in [2], [4]), permanent consumption C_t^P is hypothesized to be proportional to permanent income Y_t^P as

$$C_t^P = \beta Y_t^P \tag{5}$$

Permanent entities relate to current values according to

$$C_t = C_t^P + C_t^T \quad \text{and} \quad Y_t = Y_t^P + Y_t^T$$

with temporary components C_t^T, Y_t^T respectively. Permanent as well as temporary parts are unobservable. Following (5) we can express

$$C_t = \beta Y_t^P + C_t^T \tag{6}$$

what is a simple relation between actual consumption and permanent income. Part C_t^T represents a disturbance term. Having actual values of income, the permanent parts Y_t^P are computing following article 1. Parameter β is easy to be estimated by using OLS.

Dynamic properties of (6) can be seen after substituting (1) to (6).

$$C_t = \beta\lambda Y_t + \beta(1-\lambda) Y_{t-1}^P + C_t^T \quad . \tag{7}$$

From (6) we also have $\beta Y_t^P = C_t - C_t^T$ and hence $\beta Y_{t-1}^P = C_{t-1} - C_{t-1}^T$ and substituting in (7)

$$C_t = \beta\lambda Y_t + (1-\lambda) C_{t-1} + C_t^T - (1-\lambda) C_{t-1}^T = \beta\lambda Y_t + (1-\lambda) C_{t-1} + \theta$$

The marginal propensity to consume $\frac{d C_t}{d Y_t} = \lambda\beta$. In the long – run, equation (7) changes into

$$\overline{C} = \beta\lambda\overline{Y} + (1-\lambda)\overline{C} \quad \text{and therefore} \quad \overline{C} = \beta\overline{Y}$$

with $\overline{C}, \overline{Y}$ representing the values of long – run equilibrium. Under the assumption made about λ, it is evidently $\beta > \lambda\beta$.

By such a way, a permanency hypothesis is adopted without asking a question about its validity. A justification of a permanency assumption is possible to study using an approach of Campbell and Mankiw [1]. The idea

is to distinguish between two groups of consumers those who consume their current disposable income $C_{1t} = Y_{1t}$ and those who consume their permanent disposable income $C_{2t} = Y_{2t}^P$. Total disposable income can be seen as

$$Y_t = Y_{1t} + Y_{2t}^P = \omega Y_t + (1-\omega) Y_t .$$

Hence $C_{1t} = \omega Y_t$ and $\Delta C_{1t} = \omega \Delta Y_t$, and similarly $C_{2t} = (1-\omega) Y_t$ and $\Delta C_{2t} = (1-\omega)\Delta Y_t$. According to Flavin [3], consumption should respond to innovations in current income because these innovations provide new information about future income and therefore induce revisions in permanent income. That is why the last equation can also be expressed as $\Delta C_{2t} = \alpha + (1-\omega)\varepsilon_t$ where α is a constant and ε_t is the innovation.

In general, we can see current income as an autoregressive process

$$Y_t = \gamma_0 + \gamma_1 Y_{t-1} + \gamma_{1-2} Y_{t-2} + ... + \gamma_p Y_{t-p} + \varepsilon_t \tag{8}$$

The increment of consumption then is

$$\Delta C_t = \alpha + \beta_0 \Delta Y_t + \beta_1 \Delta Y_{t-1} + ... + \beta_p \Delta Y_{t-p} + (1-\omega)\varepsilon_t \tag{9}$$

In this equation the β_i parameters are measures of the "excess sensitivity" of consumption of current income [3].

The implication of permanent income hypothesis is $\forall i : \beta_i = 0$ and it is tested by running (8) and / or (9) with and without the restriction and forming an appropriate statistic. Relevant tests are

- Likelihood ratio, based on a comparison of restricted and unrestricted versions
- Wald test in the procedure of which only the unrestricted parameters are calculated
- Lagrange multiplier test into which only restricted results enter.

For details see e.g. [6].

For the change in aggregate consumption, current plus permanent, we now have

$$\Delta C_t = \Delta C_{1t} + \Delta C_{2t} = \alpha + \omega \Delta Y_t + (1-\omega)\varepsilon_t . \tag{10}$$

The increments, instead of levels, give a higher chance to work with stationary variables. Besides, only current values of C_t, Y_t enter the computation.

The permanent income hypothesis coincides with $H_0 : \omega = 0$ which can be tested after estimating ω. If H_0 is not rejected, consumption is a random walk what means that it is unpredictable. In case of rejection of the null hypothesis, consumption tracks income closely.

Technical problem arises through the fact that the correlation between ΔY_t and ε_t in (10) is not necessary zero. That is why the OLS method cannot be applied and an IV approach should be used. Having instruments $Z_1, Z_2, ..., Z_p$ we estimate

$$\Delta Y_t = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + ... + \gamma_p Z_p + \xi_{1t} .$$

From a more complex point of view we also can formulate

$$\Delta C_t = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + ... + \beta_p Z_p + \xi_{2t}$$

The independence of ΔC_t on ΔY_t means $\beta_i = 0,\ i = 1,..., p$. There is no objection to use lagged ΔY_{t-i} as instruments helping to construct ΔY_t. Hence, we are solving the same problem as in relation (9).

Calculations based on (10) instead on (9) comprise a zero restriction of β parameters explicitly. That is why the LM test is a most convenient one to be applied following the derivation of the test statistic in [6] we have $LM = nR^2$ where R^2 is the squared multiple correlation coefficient from the regressions of residuals of (10) on all the data including ΔY_{t-i} instruments. It is $LM = \chi^2(p),\quad p$ being the number of restrictions.

4 Application to Czech Economy

Permanent income and consumption hypothesis was applied to the Czech economy concerning 1995Q1 to 2011Q4 data about disposable income and final consumption both in Euro per inhabitant, source Eurostat. Time series ΔC_t, ΔY_t, are stationary according to the ADF test.

As the instruments to build ΔY_t its four lags were used as

$$\Delta Y_t = \gamma_0 + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \gamma_3 \Delta Y_{t-3} + \gamma_4 Z_{t-4} + \xi_{1t}$$

then the relation

$$\Delta C_t = \alpha + \omega \Delta \hat{Y}_t + (1-\omega)\varepsilon_t$$

estimated. The result $\hat{\omega} = 0.864$ with *t*-probability = 0.000 do not support the permanent income hypothesis, nevertheless the LM test was performed. The former intuition is confirmed by the finding that LM = 0.237 x 63 = 14.931; $\chi^2(4) = 14.931 > \chi^2_{crit}(4) = 0.7107$ at 5 % significance level. Similar computations were repeated with the data covering the pre-crisis period only (from beginning of 1995 to the end of 2007). The results do not differ significantly. Detailed computations in Appendix.

5 Conclusions

Consumption following a permanent income is a theoretical concept the confirmation or non-confirmation of which brings a consequence to an eventual forecast of future consumption. An econometric approach for testing the validity of permanency is described. In case of a positive answer, a way how to compute permanent income given the current one is briefly recapitulated including the short – run as well as the long – run impact. Using the actual data representing the Czech economy the permanency hypothesis is investigated and rejected. The same result appeared even after dropping the last observations coinciding with the period of financial and / or economic crises. We conclude that in the CR a consumption tracks income very closely.

Acknowledgements:

Financial support of IGA F4/1/2012 is gratefully acknowledged by the author.

References

[1] Campbell, J. Y., Mankiw, N. G.: Permanent Income, Current Income and Consumption, *Journal of Business & Economic Statistics*, Vol. 8, No. 3 (1990) pp. 265 – 279
[2] Dougherty, Ch.: *Introduction to Econometrics*, Oxford Univ. Press, 1992
[3] Flavin, M. A. The Adjustment of Consumption to Changing Expectations about Future Income, *The Journal of Political Economy*, Vol. 89, No. 5, (1981), pp. 974-1009

[4] Friedman, Milton. *A Theory of the Consumption Function*. Princeton, N.J.: Princeton Univ. Press, 1957.
[5] Hall, Robert E. "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence." *Journal of Political Economy* 86, no. 6 (1978): 971-87.
[6] Johnston, J., DiNardo, J.: *Econometric Methods*, McGraw Hill, 1997
[7] Lukáčik, M.; Szomolanyi, K.; Lukáčiková, A.; Hančlová, J.. Long-Run Structural Macroeconometric Models of the Slovak and Czech Economies. In *Proceedings of the International Scientific Conference Quantitative Methods in Economics (Multiple Criteria Decision Making XV).* Bratislava: University of Economics in Bratislava (2010) pp 107-116.
[8] Sargent, Thomas J. "Rational Expectations, Econometric Exogeneity, and Consumption." *Journal of Political Economy* 86, no. 4 (1978): 673-700.
[9] Sims, C. A., Implications of rational inattention (2002) http://pages.stern.nyu.edu/~dbackus/Exotic/1Robustness/Sims%20inattention%20JME%2003.pdf

Appendix

Using PCGive to 1995Q1 to 2011Q4 data about disposable income and final consumption both in Euro per inhabitant, source Eurostat, the results are:

	Coefficient	**Std.Error**	**t-prob**	**R^2**
Modelling DY				0.823137
DY_1	-0.323146	0.1050	0.003	
DY_2	-0.267432	0.1055	0.014	
DY_3	-0.315412	0.1051	0.004	
DY_4	0.615317	0.1076	0.000	
Constant	21.2197	10.88	0.056	
Modelling DC				0.712282
Constant	13.6836	10.44	0.195	
Fitted DY	0.864943	0.07038	0.000	
Modelling residuals				0.237802
Fitted DY	-0.499695	0.5356	0.355	
DY_1	0.158083	0.1498	0.296	
DY_2	0.0267578	0.1292	0.837	
DY_3	-0.189746	0.1486	0.837	
DY_4	0.458985	0.3993	0.255	

Causality in mean and variance between returns of crude oil and metal prices, agricultural prices and financial market prices

Monika Papież [1], Sławomir Śmiech[2]

Abstract. In the short run the rate of return on commodities mostly depends on investors' reaction to incoming information on the world economic situation, that is the global demand, as well as information on the fluctuations in the supply of those commodities. The paper presents the analysis of dependencies between the prices of crude oil and various metals, energy sources, agricultural raw materials, food and beverages and the variables specific for the financial market. The methodology was based on Cheung and Ng and Hong tests, which allow to analyse Granger causality of daily returns both in mean and variance. The results of the analysis indicated the existence of simultaneous dependencies between the prices of crude oil and the prices of other commodities. The analysis also revealed that the price of crude oil is the Granger cause of natural gas, S&P 500, coffee, corn, cotton and copper prices. The results of the analysis indicate that the prices of platinum, natural gas and the value of US 30 Year Bonds were the Granger cause of the crude oil prices. Causality of returns in variance was observed in several pairs only.

Keywords: crosscorelations, arma-egarch, causality in mean, causality in variance.

JEL Classification: C32, G15, O13, Q37
AMS Classification: 91B84, 62M10

1 Introduction

The rise in the prices of a substantial number of commodities can be observed recently. This phenomenon can be explained by the fact that certain sources react to global macroeconomic factors in a similar way. Energy, mostly obtained from crude oil and gas, is mostly used in industry. Thus, an expected increase in production and an increased demand for metals used in industry (copper, silver, platinum) should cause the reaction of the prices of energy sources. Energy prices should also react to the changes in the values of stock market indices (SP 500) or the monetary policy (US30Y Bond), which might be treated as aggregate data on the condition of the US economy and the world economy. On the other hand, key information for the crude oil supply (mainly armed conflicts on the areas belonging to oil producers) will influence metal prices and the state of the world economy. It can be expected that there will be connections between USD dollar exchange (USD Index) and the prices of commodities paid for in dollars even though produced in countries with other currencies. The links between crude oil prices and crops (coffee, cocoa, soybean, rice and cotton) are different. Kilian, Park [9] claim that the price of crude oil has the greatest influence on food prices, because an increasing price of crude oil increases both the transport costs and food production costs through the increase 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.). This means that farmers replace e.g. cotton with corn, which in turn leads to the increase of cotton prices due to its lack. Also, the demand for coffee, cocoa, rice and soybean is not flexible (consumers' habits). Coffee and cocoa are produced on the southern hemisphere, but consumed mostly on the northern hemisphere (Europe, the USA), so the prices of those products depend on the costs of transport (crude oil), unlike soybean or cotton, which are mainly produced and processed in Asia (China). Additionally, those plants are annual, which means that the changes in their prices can lead to the changes of crops; they are also sensitive to weather factors. As a result, the existence of the cycles can be expected in case of those plants (similar to so called pork cycles). Gold is a different category, as it does not play an important role in the industry and is treated as a form of capital investment.

All above mentioned commodities are traded on commodities exchanges. A significant volume of transactions is not connected with the physical delivery of commodities, but is based on the settlement of contracts (the

[1] Cracow University of Economics, Faculty of Management, Department of Statistics, Rakowicka 27, 31-510 Cracow, Poland, papiezm@uek.krakow.pl.
[2] Cracow University of Economics, Faculty of Management, Department of Statistics, Rakowicka 27, 31-510 Cracow, Poland, smiechs@uek.krakow.pl.

differences between the prices of an underlying instrument and an exercise price). Thus, for most participants commodities are a form of securing of investment portfolios. In this case, it may turn out that in the short run the prices of commodities are shaped not by fundamental factors but by broadly understood investors' strategies.

The aim of the paper is to check short term connections between a wide range of financial instruments, such as, prices of energy sources, metals, food, currencies, treasury bills and crude oil. The approach adopted in the analysis allowed for the evaluation of causality in mean (for the rate of return) and causality in variance. In the first case, the hypothesis verified states that past values of return of one instrument are correlated with current returns of another instrument. In the second case, the connections between conditional variance of instruments returns are analysed. This approach allows for the analysis of price co-movement and information transmission analysis.

The paper investigates the following research hypotheses: in the short run the rates of return of crude oil should be the Granger cause of rates of return of all above mentioned financial instruments; the rates of return of metals, treasury bills, index SP500 and natural gas are the Granger causes of crude oil; the rates of return of food are not the Granger cause of crude oil.

2 Research literature

The analysis of connections between the prices of energy sources and factors shaping the economic condition (the growth of GDP, metals prices, food prices) has been dealt with in many works. Most of them use the methodology developed for multidimensional autoregressive models (cointegration, impulse response, Granger causality). Campiche's et al. [3] findings confirmed the lack of cointegration between oil and corn, sorghum, sugar, soybeans, soybean oil, and palm oil markets. Different results appeared in Harii et al [6], who identified a long term equilibrium relationship between oil prices and all agricultural prices except wheat. Basher, Haug, Sadorsky [1] found out that positive shocks to oil prices tend to depress emerging market stock prices and the US dollar exchange rates in the short run. Nazlioglu, Soytas [10] used panel cointegration and studied dynamic relationships between the world oil prices and the prices of various agricultural commodities. The results obtained provided strong evidence of the impact of world oil price changes on agricultural commodity price. Papież, Śmiech [13] analysed the relations between crude oil prices and the prices of other energy sources, and their results confirmed a dominating role of crude oil on the primary fuels market. The dependencies between the prices of crude oil and metals were investigated by Soytas et al. [18] and Sari et al. [17] while the gas market was described by Wasilewski [19] and Rychlicki, Siemek [15]. The dynamics of the changes in the prices of raw materials can be investigated with the use of price indices, widely described by e.g. Białek [2].

3 Methodology

Most of the existing empirical studies regarding return or volatility spillover use various kinds of Granger tests in which residuals (squared residuals) of one variable are regressed by their own lags or lags of other variables. This is the way how multivariate GARCH model works. Cheung and Ng [4] proposed a two-step procedure to detect patterns of spillover between markets. The greatest advantage of this procedure is the flexible specification of the innovation process and robustness to asymmetric and leptokurtic errors. In the first step, it is necessary to estimate a time-varying conditional mean and variance models of returns in different markets,

Here we assumed that returns of each instrument price $y_{i,t}$ are characterized by the process:

$$y_{1,t} = \mu_{i,t} + z_{i,t}\sigma_{i,t} \tag{1}$$

where $\mu_{i,t}$ is the conditional mean, $\sigma_{i,t}^2$ is the conditional variance of $y_{1,t}$, $z_{i,t}$ is an independent white noise process with zero mean and unit variance. The conditional variances $\sigma_{i,t}^2$ are characterized by EGARCH models, first proposed by Nelson [11]. During the next stage, cross-correlation functions (CCF) of standardized innovations from these models are used to test causality in mean and causality in variance effects. Both kinds of causality are here interpreted as return or volatility spillover. In the paper we used modified test statistics proposed by Hong [7].

4 Data and empirical results

The analysis of causalities was conducted using the daily data (five working days per week) from the period 4 January 2006 – 30 December 2011. The data used in the analysis included the prices of futures contracts traded on the Commodity Exchange (COMEX), the New York Mercantile Exchange (NYMEX) and the Chicago Board

of Trade (CBOT) and indices. The following are the variables describing the following markets: the energy market, the financial market, the agricultural market and the metals market. The detailed description of the data set and descriptive statistics for daily time series data are presented in Table 1.

Variable	Symbol	Unit	Mean	Median	Max	Min	Std. Dev.	Skewness	Kurtosis
Crude oil	CL.F	$/bbl	79.3	76.2	144.9	34.4	20.04	0.60	3.50
Natural gas	NG.F	c/mmBtu	6.0	5.6	13.6	2.5	2.15	0.99	3.80
USD Index	USD_I	-	58.0	57.9	67.7	50.6	3.49	0.24	2.74
S&P 500	^SPX	-	1226.9	1266.3	1565.2	676.5	190.09	-0.56	2.72
US30Y Bond	US.F	-	119.5	117.9	146.1	105.1	9.05	0.90	3.31
Cocoa	CC.F	$/t	2442.0	2598.0	3748.0	1380.0	610.00	-0.25	1.78
Coffee	KC.F	c/lb	151.2	132.2	305.6	94.0	52.58	1.30	3.38
Corn	C.F	¢/bu	444.5	394.6	786.0	204.0	148.77	0.55	2.29
Soybean	S.F	¢/bu	1012.3	987.0	1635.0	537.0	269.68	0.00	2.05
Cotton	CT.F	c/lb	76.9	66.0	218.6	40.1	34.69	1.90	6.38
Copper	HG.F	¢/lb	320.5	332.5	463.1	127.9	72.78	-0.61	3.01
Gold	GC.F	$/ozt	996.2	921.4	1903.3	521.3	339.65	0.69	2.52
Platinum	PL.F	$/ozt	1433.2	1410.1	2258.3	765.2	310.65	0.26	2.32
Silver	SI.F	$/ozt	1843.6	1521.0	4846.0	884.0	862.20	1.41	3.95

Table 1 Summary statistics for daily time series

Specifically, the returns are defined as $r_t = \ln(p_t / p_{t-1})$ where p_t is the opening price on day *t*. Next, it needed to be checked if the analysed variables contain a structural break which should by taken into account. It is shown in Rodrigues et al. [14], that the procedures presented in Cheung and Ng and Hong are not resistant if there is a structural break in the volatility. Therefore, we examined the return series for the presence of a structural break in unconditional variance. We used the test proposed first by Inclan, Tiao [8] and then improved by Sanso et al., [16]. The effect of a structural brake was eliminated by the method proposed by Nouira et al. [12]. The method suggested by Nouira et al. [12] was used to eliminate the effect of a structural break in variance and to obtain filtered return series. Further analysis was conducted on these filtered series. Next, using the augmented Dickey–Fuller (ADF) method and KPSS test, it was verified that all return prices rejected the null hypothesis of the existence of a unit root, which means that the considered return series were stationary.

Because the analysis of causality in these tests was conducted on standardized residuals of ARMA-EGARCH, at first this kind of models was estimate. The models were chosen on the basis of the Akaike's information criterion. Due to lack of space, the paper contains only the classes of models without the estimated values of parameters for particular financial instruments. EGARCH (1,1) model with the Student-t distribution was estimated for cocoa, corn, soybean, cotton, copper, gold, platinum, and silver; EGARCH (1,1) model with the normal distribution was estimated for the USD Index; AR (1) – EGARCH (1,1) model with the Student-t distribution for crude oil; GARCH (1,1) model with the Student-t distribution for natural gas and coffee; GARCH (1,1) model with the normal distribution for US30Y bond; and GARCH (1,2) model with the normal distribution for S&P 500. For the estimated models the values of Box-Pierce statistics for the first 20 autocorrelation of the standardized residuals and the squared standardized residuals were statistically insignificant.

	Natural gas	**USD Index**	**S&P 500**	**US30Y Bond**	**Cocoa**	**Coffee**	**Corn**	**Soybean**	**Cotton**	**Copper**	**Gold**	**Platinum**	**Silver**
Crude oil	**0.26**	**-0.31**	**0.33**	**-0.21**	**0.22**	**0.27**	**0.17**	**0.21**	**0.22**	**0.44**	**0.40**	**0.40**	**0.45**

Table 2 Correlations between standardized residuals of crude oil and standardized residuals of other commodities and financial investments

Note: All correlation coefficients are statistically significant at the 1% level, taking into account the effect of multiple testing.

The values of correlation coefficients between standardized residuals are presented in Table 2. These values may be interpreted as contemporaneous causality between the returns on the market). The strongest positive

correlation can be observed between the standardized residuals of crude oil price and metals prices (about 0.40-0.45). This indicates a large contemporaneous causality between the prices of the crude oil and metals (copper, gold, platinum, silver) in the years 2006-2011. A significant positive correlation coefficient can also be observed between standardized residuals of crude oil price and S&P 500. However, an average negative correlation exists between the prices of crude oil and the remaining variables characteristic for the financial market. This indicates an average contemporaneous causality between the prices of the crude oil and the USD index and US 30 year bond. A significant positive correlation can be seen between standardized residuals of oil price and the prices of agricultural market (cocoa, coffee, corn, soybean and cotton). Contemporaneous causality between the prices means that information is absorbed by both markets at the same time (within the same calendar day).

Further analysis will be aimed at checking the existence of Granger causality between the returns of crude oil and other commodities and financial investments. In order to test causality in mean, standardized residuals are used from ARMA-EGARCH models and cross-correlation functions were estimated. Test Q 1 was used in the analysis of causality in mean. The value of the test was estimated for one day (M = 1), for two days (M = 2), and for one week (M = 5). The construction of Cheung and Ng test, especially the hypotheses tested, resulted in a situation in which rejecting the null hypothesis for low M values causes the rejection of the null hypothesis also for higher M values. The results of the test for the null hypothesis (crude oil is not the Granger cause in mean of other commodities and financial investments) and the results of the test for the null hypothesis (other commodities and financial investments are not the Granger cause in mean of crude oil) are given in Table 3.

	Natural gas	**USD Index**	**S&P 500**	**US30Y Bond**	**Cocoa**	**Coffee**	**Corn**	**Soybean**	**Cotton**	**Copper**	**Gold**	**Platinum**	**Silver**
Null hypothesis: the crude oil price is not the Granger cause in mean of other commodities prices													
M=1	**0.001**	0.305	**0.046**	0.196	0.649	**0.006**	**0.000**	0.243	**0.000**	**0.002**	0.464	0.736	0.773
M=2	**0.000**	0.485	**0.017**	0.317	0.713	**0.000**	**0.000**	0.399	**0.000**	**0.000**	0.628	0.735	0.742
M=5	**0.000**	0.319	**0.050**	0.236	0.758	**0.000**	**0.000**	0.520	**0.000**	**0.000**	0.744	0.797	0.820
Null hypothesis: other commodities prices are not the Granger cause in mean of the crude oil price													
M=1	**0.019**	0.732	0.101	**0.027**	0.132	**0.000**	**0.000**	**0.033**	**0.000**	0.108	0.639	**0.000**	0.859
M=2	**0.001**	0.734	0.116	**0.003**	0.185	**0.000**	**0.000**	**0.006**	**0.000**	0.131	0.709	**0.000**	0.753
M=5	**0.001**	0.686	**0.031**	**0.002**	0.334	**0.000**	**0.000**	**0.024**	**0.000**	0.119	0.343	**0.000**	**0.028**

Table 3 Causality-in-mean test results in the years 2006-2011- *p*-value are given

The results of the causality-in-mean test show that in the period 2006-2011 the price of crude oil was not the Granger cause of USD index, cocoa, gold and silver and that those prices and the index were not the Granger cause of crude oil. This means that past information regarding the price of crude oil did not improve the forecasts of the price of cocoa, gold, silver and USD index (and vice versa). However, the price of crude oil was the Granger cause of natural gas, coffee, corn and cotton prices (and vice versa), which means that the price of crude oil improved the forecasts of average prices on the energy market and the prices on the agricultural market (and vice versa). The results of the analysis indicate that the price of crude oil was the Granger cause of the price of copper and index of S&P 500 (but those two variables were not the Granger cause of the price of crude oil). Thus, the forecast of the price of Copper and the value of index S&P 500 were improved by the values of the price of crude oil. A different situation could be observed in case of the prices of platinum, soybean and the value of US 30-year bonds. They were the Granger cause of the price of crude oil but the price of crude oil was not the Granger cause of those commodities.

Cheung and Ng [4] argued that the results obtained from causality tests in variance between two different markets (variables) are affected when there is evidence of causality in mean. Therefore, as Gebka, Serwa [5] suggest, ARMA-EGARCH were re-estimated to include the lagged return of series, which were the Granger cause (causality in mean) of a given variable. Re-estimation of the models allowed to eliminate the influence of causality in mean on the values of causality in variance test. Table 4 presents the values of correlation coefficients between the squared standardized residuals. The values of those coefficients can be interpreted as contemporaneous causality between the volatility of prices and indices. The greatest positive correlation can be seen between the squared standardized residuals of crude oil and copper (0.28) and crude oil and S&P 500 (0.28). This indicates large simultaneous links between the volatility of crude oil and copper as well as crude oil and financial markets. Also, contemporaneous causality could be observed between the volatility of the crude oil price and the volatility of other metals prices (gold, platinum, silver) and the volatility of the financial market (USD index, US 30-year bonds). However, there was no contemporaneous causality between the volatility of

crude oil prices and the volatility of agricultural market, except the volatility of coffee prices.

	Natural gas	**USD Index**	**S&P 500**	**US30Y Bond**	**Cocoa**	**Coffee**	**Corn**	**Soybean**	**Cotton**	**Copper**	**Gold**	**Platinum**	**Silver**
Crude oil	**0.16**	**0.15**	**0.28**	**0.08**	0.02	**0.10**	0.04	0.04	0.04	**0.28**	**0.20**	**0.18**	**0.19**

Table 4 Correlations between squared standardized residual of crude oil and squared standardized residuals of other commodities and financial investments

Similarly to the procedure described above, test Q1 was used in the analysis of causality in variance. The value of the test was estimated for one day (M = 1), for two days (M = 2), and for one week (M = 5). The results of the test for the null hypothesis (crude oil is not the Granger cause in variance of other commodities and financial investments) and the results of the test for the null hypothesis (other commodities and financial investments are not the Granger cause in variance of crude oil) are given in Table 5. The results of the causality-in-variance test indicate that in the period in question the volatility of crude oil prices was the Granger cause of the volatility of S&P500, the volatility of silver price and volatility of gold price after the period of one week (M= 5). This means that the volatility of crude oil prices influenced only the volatility of S&P 500 and the price of silver. On the other hand, the volatility of coffee price and the volatility of copper price were the Granger cause of the volatility of crude oil price.

	Natural gas	**USD Index**	**S&P 500**	**US30Y Bond**	**Cocoa**	**Coffee**	**Corn**	**Soybean**	**Cotton**	**Copper**	**Gold**	**Platinum**	**Silver**
Null hypothesis: the crude oil price is not Granger cause in variance of other commodities prices													
M=1	0.401	0.681	**0.001**	0.061	0.882	0.342	0.283	0.927	0.767	0.422	0.188	0.255	**0.021**
M=2	0.582	0.722	**0.000**	**0.038**	0.755	0.527	0.457	0.758	0.741	0.599	0.302	0.417	**0.001**
M=5	0.745	0.222	**0.000**	0.075	0.079	0.507	0.659	0.866	0.766	0.743	**0.007**	0.202	**0.014**
Null hypothesis: other commodities prices are not the Granger cause in variance of crude oil prices													
M=1	0.387	0.545	0.839	0.532	0.173	**0.000**	0.770	0.480	0.578	**0.048**	0.716	0.764	0.570
M=2	0.570	0.673	0.751	0.667	0.273	**0.000**	0.741	0.638	0.687	**0.019**	0.730	0.740	0.684
M=5	0.708	0.100	0.313	0.684	0.456	**0.000**	0.166	0.782	0.388	0.081	0.807	0.757	0.773

Table 5 Causality -in-variance test results in the years 2006-2011**-** *p*-value are given

5 Conclusions

The analysis of Granger causality in mean and variance between the price of crude oil and the prices of other commodities allowed to verify the hypothesis stated at the beginning of the paper. The analysis indicated simultaneous dependencies between the prices of crude oil and the prices of other commodities. This simultaneous causality may indicate that all analysed prices are influenced by similar fundamental factors. Unfortunately, on the basis of the analyses conducted it is not possible to answer the question whether it is the information on the world conjecture or the information on disturbances in supply. It is worth noticing that a significant negative correlation could be observed only between the prices of crude oil and the USD index and US 30 year bonds. The analysis indicated that contemporaneous causality was observed between the volatility of the crude oil price and the volatility of metals prices, and the volatility of financial market, but there were no simultaneous dependencies between the volatility of crude oil price and the volatility of the agricultural market, except the volatility of coffee prices. Unfortunately, the hypothesis stating that in the short run the rates of return of crude oil are the Granger cause of the rates of return of other commodities was not confirmed. The results of the causality-in-mean tests show that the price of crude oil was the Granger cause of the prices of natural gas, S&P 500, coffee, corn, cotton and copper. Similarly, it cannot be generalised that the rates of return of metals, treasury bills, index SP500 and natural gas were the Granger causes of crude oil. The results of the analysis indicate that the prices of platinum, natural gas and the value of US 30-year bonds were the Granger cause of the crude oil price. Also, the rates of return of food (except cocoa) were the Granger cause of crude oil.

References

[1] Basher, S. A., Haug, A. A., Sadorsky, P.: Oil prices, exchange rates and emerging stock markets. *Energy Economics* **34** (2012), 227-240.

[2] Białek, J.: Proposition of a General Formula for Price Indices. *Communications in Statistics -Theory and Method,* **41** (2012), 943-952.

[3] Campiche, J. L., Bryant, H. L., Richardson, J. W., Outlaw, J. L.: Examining the evolving correspondence between petroleum prices and agricultural commodity prices. *The American Agricultural Economics Association Annual Meeting*, Portland, OR, July 29-August 1, 2007.

[4] Cheung, Y. W., Ng, L. K.: A causality-in-variance test and its application to financial market prices. *Journal of Econometrics* **72** (1996), 33–48.

[5] Gebka, B., Serwa, D.: Intra- and inter-regional spillovers between emerging capital markets around the world. *Research in International Business and Finance* **21** (2007), 203-221.

[6] Harri, A., Nalley, L., Hudson, D.: The relationship between oil, exchange rates, and commodity prices. *Journal of Agricultural and Applied Economics* **41** (2009), 501–510.

[7] Hong, Y.: A test for volatility spillover with application to exchange rates. *Journal of Econometrics,* **103** (2001), 183–224.

[8] Inclan, C., Tiao, G. C.: Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association* **89** (1994), 913-923.

[9] Kilian, L., Park, C.: The impact of oil price shocks on the U.S. stock market. *International Economic Review* **50** (2009), 1267-1287.

[10] Nazlioglu, S., Soytas, U.: Oil price, agricultural commodity prices, and the dollar: A panel cointegration and causality analysis, *Energy Economics* **34** (2012), 1098–1104.

[11] Nelson, D. B.: Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* **59** (1991), 347—370.

[12] Nouira, L., Ahamada, I., Jouini, J., Nurbel, A.: Long-memory and shifts in the unconditional variance in the exchange rate Euro/US Dollar returns. *Applied Economics Letters* **11** (2004), 591-594.

[13] Papież, M., Śmiech, S.: The analysis of relations between primary fuel prices on the European market in the period 2001-2011. *Rynek Energii* **5(96)** (2011), 139-144.

[14] Rodrigues, P. M. M., Rubia A.: Testing for causality in variance under nonstationarity in variance. *Economics Letters* **97** (2007), 133-137.

[15] Rychlicki, S., Siemek, J.: Energy projections for thee world with particular focus on natural gas by the International Gas Union. *Rynek Energii* **6(91)** (2010), 3-7.

[16] Sansó, A., Aragó, V., Carrion, J. Ll.: Testing for changes in the unconditional variance of financial time series. *Revista de Economía Financiera* **4** (2004), 32-53.

[17] Sari, R., Hammoudeh, S., Soytas, U.: Dynamics of oil price, precious metal prices, and exchange rate. *Energy Economics* **32** (2010), 351–362.

[18] Soytas, U., Sari, R., Hammodeh, S., Hacihasanoglu, E.: World oil prices, precious metal prices and macroeconomy in Turkey. *Energy Policy* **37** (2009), 5557–5566.

[19] Wasilewski, A.: Economic global crisis and natural gas. *Rynek Energii* **3(88)** (2010), 3-7.

Skip pickup and delivery problem with vehicles circulation

Jan Pelikán[1]

Abstract. The skip transport consists in transport of skips (big containers, trailers) from initial location to destination location using vehicles (tractors). Capacity of vehicles is limited, usually the capacity of the vehicles is one or two containers. The problem is defined on the graph, nodes are initial and destination locations, total transport cost is minimized. It is supposed that depot is given, all vehicles start and finish in the depot. There are many papers which pay attention to skip delivery problem (SDP). It was shown that the problem SDP can be solved as b-matching problem for case of vehicle capacity two, and the case of capacity one or two. Unlike literature results the skip pickup and delivery problem with vehicle circulation is studied (SPDPC). Depots of vehicles are not given, but for each vehicle the cyclical path in the graph should be found and depot the path can be arbitrary node on this path. All cyclical paths create a circulation of the multigraph, where each arc can be multiplied. The total transport costs have to be minimized. The problem SPDPC is a special case of the pickup and delivery problem with transfers and split demand, but the transport demand, the capacity of containers and a solution of the problem are integer. A mathematical model is formulated for the case of distance matrix with triangular inequality and the capacity of vehicles one. The matrix of the model is totally unimodular then the problem is polynomially solved. A method for the the circulation of vehicles is proposed. It is demonstrated on a numerical example.

Keywords: integer programming, skip pickup and delivery problem, logistic models

JEL classification: C44
AMS classification: 90C15

1 Introduction - skip delivery problem

There are many papers that pay attention to skip delivery problem. In [1] the transport problem of skips is studied, in which skips are transported by vehicles with capacity one or two from depot to customers. There is a set of customers with the skip demand. Each customer is the node in the graph $G = \{V, E\}$, V is a set of nodes, E a set of edges, node 1 is depot. Demand of node i is denoted by positive integer b_i. The cost of travel from node i to node j is c_{ij}, it is supposed that the triangular inequality is satisfied and the matrix C is symmetrical. Number of vehicles in the depot is unlimited, the capacity Q of vehicles is fixed and positive integer and the depot of all vehicles is the node 1. All skips are available in the depot and from the depot are delivered to the nodes according the node demand b_i.

1.1 Skip delivery problem [1], case Q=2

In the case Q=2 there are only two possible round trips:
a) from the depot 1 to the node i and the node j and return to the node 1, this trip we denote e_{ij},
b) from the depot 1 to the node i and return to the depot, the trip is e_{ii}.

Now we define an artificial graph problem: minimum weight b-matching problem as follows. The graph $G' = \{V', E'\}$, where $V' = V - \{1\}$. The set of edges E' contains all round trips e_{ij} and e_{ii},

[1]University of Economics Prague/W. Churchill Sq.4, Prague, e-mail: pelikan@vse.cz

where trip e_{ij} is edge of the graph G' and e_{ii} the loop of the graph G'. The weight of the edge e_{ij} is $c'_{ij} = c_{1i} + c_{ij} + c_{j1}$, and the weight of the loop e_{ii} is $c_{ii} = c_{1i} + c_{i1}$. SDP problem with Q=2 can be formulated as minimal weighted b-matching problem of the graph G'. The mathematical formulation of the b-matching problem is as follows: Let us have integer variables x_{ij}, where (i, j) is edge or loop of the graph G'. The value x_{ij} for $i \neq j$ is a number of trips of the type a) and x_{ii} a number of trips of the type b). Mathematical model of the minimal weighted b-matching problem is:

$$z = \sum_{(i,j)\in E'} c'_{ij} x_{ij} \to min \tag{1}$$

$$\sum_{(i,j)\in E', i\neq j} x_{i,j} + \sum_{i\in V'} 2x_{ii} \geq b_i, i \in V' \tag{2}$$

$$x_{i,j}, \text{integer}, (i,j) \in E' \tag{3}$$

The problem is reducible, i.e. there is an optimal solution of the instance in which each node is served by as many full load depot-node trips as possible. Reducible instance can be transformed into an instance of the generalized minimum cost matching problem, because demand of nodes is at least only one container.

1.2 Skip delivery problem [1], case Q=1 or 2

Similar results hold for the case with capacity of vehicles is one or two. At first the cost of transfer thru the edge (i, j) must differ for vehicle with capacity one or capacity two. The cost of travel from node i to node j is $c'_{ij} = k_1 c_{ij}$ if the capacity of the vehicle is one, and $c'_{ij} = k_2 c_{ij}$ if the capacity is two, where c_{ij} is distance from node i to node j. If $k_2/k_1 \leq 1$ we will use only vehicle with capacity two, in case $k_2/k_1 \geq 2$ all used vehicles are with capacity one in the optimal solution. In the case Q=1 or 2 we construct the undirected complete graph $G' = \{V, E'\}$ as an instance of the minimum weight b-matching problem in the following way: $c'_{ij} = k_2(c_{1i} + c_{ij} + c_{j1}), i \neq j, i, j \in V - \{1\}$ and x_{ij} gives the number of trips $1 - i - j - 1$ with a vehicle with capacity two, $c'_{1i} = 2k_1 c_{1i}$ for $i \in V - \{1\}$ and x_{1i} is the number of trips $1 - i - 1$ and vehicle capacity one, $c'_{ii} = 2k_2 c_{1i}$ for $i \in V - \{1\}$ and x_{ii} is number of trips $1 - i - 1$ and vehicle capacity two.

Example. Given graph with 4 nodes $V = 1, 2, 3, 4$, node 1 is depot. Skip demand is $b = (0, 3, 5, 4)$, , cost matrix is $C = \begin{pmatrix} 0 & 15 & 25 & 31 \\ 15 & 0 & 19 & 27 \\ 25 & 19 & 0 & 29 \\ 31 & 27 & 29 & 0 \end{pmatrix}$.

The mathematical model of the case $Q = 2$ is:

$$z = 30x_{22} + 59x_{23} + 73x_{24} + 50x_{33} + 85x_{34} + 62x_{44} \to min$$

$$2x_{22} + x_{23} + x_{24} \geq 3$$

$$x_{23} + 2x_{33} + x_{34} \geq 5$$

$$x_{24} + x_{34} + 2x_{44} \geq 4$$

$$x_{i,j}, \text{integer}, i, j = 2, 3, 4.$$

The optimal solution is $x_{22} = 1, x_{23} = 1, x_{33} = 2$ with $z = 313$.

The mathematical model of the case $Q = 1, 2$ and $k_1 = 4/5, k_2 = 5/4$ is:

$$z = 24x_{12} + 40x_{13} + 49x_{14} + 30x_{22} + 59x_{23} + 73x_{24} + 50x_{33} + 85x_{34} + 62x_{44} \to min$$

$$x_{12} + 2x_{22} + x_{23} + x_{24} \geq 3$$

$$x_{13} + x_{23} + 2x_{33} + x_{34} \geq 5$$

$$x_{14} + x_{24} + x_{34} + 2x_{44} \geq 4$$

$$x_{i,j}, \text{integer}, i, j = 1, 2, 3, 4.$$

The optimal solution is $x_{14} = 4, x_{22} = 1, x_{23} = 1, x_{33} = 2$ with $z = 285$.

Under assumption of symmetrical cost matrix and holding triangular inequality the problem with capacity vehicles one or two is reducible.

The b-matching problem can be solved in polynomial time (see [1][2]) because the polytope of the b-matching model (1) (2) (3) with the blossom inequality (see [4]) is integral.

2 Skip Pickup and Delivery Problem

2.1 Definition of skip pickup and delivery problem

Pickup and delivery problem is defined in [5]. Given a distribution network with a set of n nodes and the cost matrix C of the travel costs between all pairs of nodes, where c_{ij} is the cost - distance between nodes i and j. Let us denote d_{kl} the number of skips that has to be transported from node k to node l. Vehicles with capacity Q are used for pickup and delivery and they can start in any node. All routes have to be cyclical, each vehicle has to come back to the node it starts from. The objective is to minimize total cost of all routes. The optimal solution is a set of cyclical routes, for each of them a depot is specified. The cyclical routes have to cover all pickup and delivery demands D. Triangular inequality and symmetry for C is supposed. The problem can be solved by the method which is shown on the example as follows.

The skip pickup and delivery problem is a special case of the pickup and delivery problem with transfers and split demand SDPDPT [3], where capacity of vehicles is Q=1. The mathematical model of SDPDPT is:

Parameters:

C	matrix of the length of arcs,
D	matrix of the goods flows between two nodes,
$Q = 1$	the capacity of the vehicle,
n	the number of nodes.

Variables:

y_{ij}	number of vehicles going thru the arc (i, j) $(i, j = 1, 2, ..., n + 1,\ i \neq j)$,
x_{ij}^{kl}	an amount of goods (a part of the total amount d_{kl}) transported from node i to node j, $(i, j = 1, 2, ..., n + 1,\ i \neq j,\ k, l = 1, 2, ..., n,\ k \neq l)$.

The SDPDPT model:

$$F(Y) = \sum_{i,j} c_{ij} y_{ij} \rightarrow min, \tag{4}$$

$$\sum_i y_{ij} - \sum_i y_{ji} = 0,\ j = 1, 2, ..., n \tag{5}$$

$$\sum_i x_{ij}^{kl} - \sum_i x_{ji}^{kl} = \left\{ \begin{array}{ll} -d_{kl}, & j = k \\ d_{kl}, & j = l \\ 0, & j \neq k, l \end{array} \right\} k, l = 1, 2, ..., n,\ k \neq l \tag{6}$$

$$\sum_{k,l} x_{ij}^{kl} \leq Q y_{ij},\ i, j = 1, 2, ..., n,\ i \neq j \tag{7}$$

$$x_{ij}^{kl} \geq 0 \text{ integer},\ k, l = 1, 2, ..., n,\ k \neq l,\ i, j = 1, 2, ..., n,\ i \neq j \tag{8}$$

$$y_{ij} \geq 0 \text{ integer},\ i, j = 1, 2, ..., n,\ i \neq j\,. \tag{9}$$

Comment.The model (5)-(10) is a multi-product flow problem, then the matrix of the constraints is not totally unimodular, even if $Q = 1$, and (see [4] [2]) therefore the polyhedron of the model is not integral.

Lemma 1: If the non negative costs matrix satisfies the triangular inequality then the length c_{ij} of the arc (i, j) is less or equal than the length of any path from the node i to the node j.

Proof. Let us have the path $(k_1, k_2, ..., k_s)$, where $k_1 = i$ and $k_s = j$ then $c_{ij} \leq c_{i,k_2} + c_{k_2,j} \leq c_{k_1,k_2} + c_{k_2,k_s} \leq c_{k_1,k_2} + c_{k_2,k_3} + c_{k_3,k_s} \leq \cdots \leq c_{k_1,k_2} + c_{k_2,k_3} + \cdots + c_{k_{s-1},k_s}$.

Lemma 2: Let (Y, X) holds the constraints (6)-(10) and one skip from the transport requirement d_{ij} is transported thru the path $P = (k_1, k_2, ..., k_s)$, where $k_1 = i$ and $k_s = j$. Let (Y', X') is the solution obtained from (Y, X), where the transport of this skip thru the path P is replaced by the arc (i, j). Then $F(Y) \geq F(Y')$.

Proof follows from the lemma 1.

Proposition 1. *Let the matrix C is nonnegative symmetric with triangular inequality, transport requirement matrix D is integer and the capacity of vehicles is equal one. Then the skip pickup and delivery problem can be solved in polynomial time by the model (11)-(13).*

Proof. It follows from lemma 2 that all skip transport requirements d_{ij} have to transported direct thru arc (i, j). So the solution Y has to meet the flow equation (12) to ensure the existence a set of cyclical routes and inequalities (13) which ensures all transport requirements D. The constraints matrix of the model (11)-(13) is node-arc matrix, which is totally unimodular, therefore the polytope of the model is integral, so all extremal solutions are integer. □

$$z = \sum_{(i,j)} c_{ij} y_{ij} \rightarrow min \tag{10}$$

$$\sum_{(i,j),i \neq j} y_{ij} + \sum_{(j,k),j \neq k} y_{jk} = 0,\ j = 1, 2, ..., n \tag{11}$$

$$y_{ij} \geq d_{ij},\ i, j = 1, 2, ..., n \tag{12}$$

Example. There are 5 nodes and capacity of vehicle one. Travel costs C and transport requirements Q are :

$$C = \begin{pmatrix} 0 & 3 & 7 & 3 & 2 \\ 3 & 0 & 2 & 1 & 6 \\ 7 & 2 & 0 & 2 & 3 \\ 3 & 1 & 2 & 0 & 6 \\ 4 & 6 & 3 & 6 & 0 \end{pmatrix} \quad D = \begin{pmatrix} 0 & 0 & 0 & 0 & 4 \\ 1 & 0 & 2 & 5 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \end{pmatrix}$$ and the solution $$Y = \begin{pmatrix} 0 & 0 & 0 & 0 & 4 \\ 1 & 0 & 2 & 5 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 4 & 0 & 0 & 0 \\ 2 & 3 & 0 & 0 & 0 \end{pmatrix}$$ is optimal.

The number of vehicles traveling thru edge (i, j) without load is $y_{32} = 1, y_{41} = 3, y_{42} = 4, y_{51} = 2$ with cost 17. The number of full loaded vehicle going thru edges is $y_{ij} = q_{ij}$ with costs 41. Total costs 41+17=58 are optimal.

2.2 Cyclical routes generation

A number of vehicles entering each node equals to a number of vehicles leaving it in the optimal solution of the model (11), (12) and (13). For generation of a set of the cyclical routes, each in the form of a path $(i_1, i_2, ..., i_t)$, the following general algorithm can be used (see [3]):

Algorithm for the route generation:
Step 1. If $y_{ij} = 0$ for all arcs (i, j), it is not possible to generate any route, otherwise select any arc $(i_1, i_2), y_{i_1,i_2} > 0$. Set $y_{i_1,i_2} = y_{i_1,i_2} - 1$ and $t = 2$.

Step 2. Repeat while $i_1 \neq i_t$: Select any arc (i_t, y_{t+1}) for $y_{i_t,i_{t+1}} > 0$. Set $y_{i_t,i_{t+1}} = y_{i_t,i_{t+1}} - 1$ and $t = t + 1$.

Example.

Using the optimal value of variables y and the algorithm for cyclical route generation we get the routes in the table 1.

cyclical route	number of routes
1-5-1	2
1-5-2-1	1
1-5-2-4-1	1
2-4-2	4
2-3-5-2	1
2-3-2	1

Table 1: Optimal routes

Acknowledgements

This research was supported by GAČR grant No. GACR P403/12/1947 and the project F4/18/2011 founded by the Internal Grant Agency of the University of Economics, Prague.

References

[1] Archetti, C., Mansini, R, Speranza, M.G.: Complexity and reducibility of the skip delivery problem, *Transportation Science* **39** (2005), pp. 182-187.

[2] Černý, M.: Binary Segmention and Bonferroni-Type Bounds. In: *Kybernetika* **47** (2011), No. 1, pp. 38-49. ISSN 0023-5954.

[3] Fábry, J.: Optimization of routes in pickup and delivery problem. In: *Mathematical Methods in Economics. České Budějovice : University of South Bohemia*, pp. 128-131. ISBN 978-80-7394-218-2.

[4] Korte, B., Vygen, J.: *Combinatorial Optimization.* Springer, New York, 2002.

[5] Savelsbergh, M., W., P., Sol, M.: *The general pickup and delivery problem.* Transportation Research, No. 29, pp. 17-29.

Double system parts optimization: static and dynamic model

Jan Pelikán[1], Jiří Henzler[2]

Abstract. A proposed optimization model deals with the problem of reserves for the functional components-parts of mechanism in order to increase its reliability. The following factors are taken into consideration: the probability of the failure-free run of a part without a reserve, the probability of the failure-free run of a part with a reserve, the mean value of losses caused by the part's malfunction without a reserve, the mean value of losses caused by the part's malfunction with a reserve, costs of the purchase and maintenance of the reserve for the given parts. The values of these parts' failure probabilities are supposed to be known in advance, the losses caused by this failure are estimated. Statistical independence of the failures of those parts is supposed. In the model, the costs of the parts' doubling are supposed to be limited to a fixed value. As a result of the problem solution, the parts of the model are sorted into two groups: the parts which are to be doubled and the parts which are not. The static model as well as the dynamic one, in which the failures are considered as Poisson events are described, including numerical examples.

Keywords: doubling of system parts, optimization model, probability of failure, mean value of losses, system reliability, static and dynamic model.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Complex mechanisms are composed of a great number of components. Each of the parts is responsible for the right functioning of the whole system and vice versa, each part's failure can disturb the system or completely put it out of operation and cause damages in its effect.

One of the ways of eliminating or at least diminishing these damages is doubling of some important parts. Having these parts doubled, there is a possibility to exchange immediately a non-functional part by a functional one (or in other words, the failure is reduced only to a necessary time of a switch-over).

On the other side there are costs of doubling, primarily the price of a doubled part. For that reason not all of the parts can be doubled, especially the expensive ones and also those whose failure is not bringing so expensive damages.

2 Double system parts optimization model

If we want to know which parts are to be doubled, the following optimization model can be used. First we introduce the presumption of the model.

Let us suppose n parts of the system (aggregates, components) $Z_1, Z_2, \ldots, Z_n$. Each of these parts is characterized by:

p_i and $\overline{p}_i$ – probabilities of the failure-free run of Z_i without and a with reserve,

q_i and $\overline{q}_i$ – the mean value of losses caused by Z_i's disorders without and with a reserve,

[1]University of Economics Prague, Departement of Econometrics, nám W. Churchilla 4, 130 67, Praha 3, Czech Republic, pelikan@vse.cz
[2]University of Economics Prague, Departement of Mathematics, Ekonomická 957, 148 00, Praha 4, Czech Republic, henzler@vse.cz

c_i – costs of the purchase and maintenance of the reserve for Z_i.

Next we suppose:

- a statistical independence of the failures of those parts,
- the costs of the parts' doubling are limited to the amount K.

Let us introduce 0-1 variables $x_1, x_2, \ldots, x_n$, the variable x_i involves the decision between the doubling of Z_i ($x_i = 1$) or not ($x_i = 0$), see [3]. Total costs of the reserves for the parts are $\sum_{i=1}^{n} c_i x_i$ and since the resources for reserves are limited by K, so it has to be valid

$$\sum_{i=1}^{n} c_i x_i \leq K. \tag{1}$$

On that conditions we can:

a) maximize the reliability of the system, that is failure-free run,

b) minimize the mean value of the sum of losses caused by the part's disorders.

In the case a) the probability of the failure-free state of the system is the product of the probabilities of the failure-free states of all the parts.

The part Z_i will be failure-free with the probability $\overline{p}_i$, if it has a reserve ($x_i = 1$). If the part Z_i is without reserve ($x_i = 0$) then the failure-free probability is p_i. Altogether the probability of the part Z_i's failure-free state can be put in the form $p_i + (\overline{p}_i - p_i)x_i$ (see [1]). Hence the total probability of the failure-free state expresses the whole system's reliability:

$$r(x) = \prod_{i=1}^{n} [p_i + (\overline{p}_i - p_i)x_i]. \tag{2}$$

After logarithming the function $r(x)$ in order to make the objective function linear we get the objective function in the form

$$\log r(x) = \sum_{i=1}^{n} \log[p_i + (\overline{p}_i - p_i)x_i]. \tag{3}$$

Since the expression $\log[p_i + (\overline{p}_i - p_i)x_i]$ for $x_i = 0$ equals $\log(p_i)$ and for $x_i = 1$ equals $\log(\overline{p}_i)$, we can write the expression $\log[p_i + (\overline{p}_i - p_i)x_i]$ in the form $(1 - x_i)\log(p_i) + x_i \log(\overline{p}_i)$":

$$\log r(x) = \sum_{i=1}^{n} [(1 - x_i)\log(p_i) + x_i \log(\overline{p}_i)] = \sum_{i=1}^{n} \log(p_i) + \sum_{i=1}^{n} x_i \log(\overline{p}_i / p_i) \tag{4}$$

The maximum reliability model is as follows:

$$\log r(x) = \sum_{i=1}^{n} [\log(p_i) + x_i \log(\overline{p}_i / p_i)], \qquad \sum_{i=1}^{n} c_i x_i \leq K, \quad x_i \in \{0, 1\}, \quad i = 1, 2, ..., n. \tag{5}$$

In the case b) the mean value of the losses caused by the part Z_i's failure is without the reserve q_i and with the reserve $\overline{q}_i$. The mean value of the total losses is as follows:

$$z(x) = \sum_{i=1}^{n} [(1 - x_i)q_i + x_i \overline{q}_i] = \sum_{i=1}^{n} [q_i - x_i \Delta q_i], \qquad \text{where } \Delta q_i = q_i - \overline{q}_i. \tag{6}$$

The minimal losses model is as follows:

$$z(x) = \sum_{i=1}^{n} [q_i - x_i \Delta q_i] \longrightarrow \min, \qquad \sum_{i=1}^{n} c_i x_i \leq K, \quad x_i \in \{0, 1\}, \quad i = 1, 2, ..., n. \tag{7}$$

3 Two models

Now, two models can be distinguished. The first one is a one-case situation, when the function operates for a short period - static case. The second one is a long-time model for a longer time period - dynamic model.

3.1 Static model

The probability of the components' failure Z_i will be denoted as π_i. Consequently the probability of the failure-free run of the part without a reserve is $p_i = 1 - \pi_i$ and the probability of the failure-free run of the part with a reserve is $\overline{p}_i = 1 - \pi_i^2$.

Let the loss caused by one failure of the part Z_i be denoted by Q_i; then the mean value of the loss is

$q_i = \pi_i Q_i$ in case when there is no reserve,

$\overline{q_i} = \pi_i^2 Q_i$ in case when there is a reserve for the part Z_i.

Example 1. Let us have parts Z_1, Z_2, Z_3, Z_4, Z_5. Main characteristics of those parts are contained in Table 1. The costs of the parts' doubling are limited by the amount $K = 100$.

	Z_1	Z_2	Z_3	Z_4	Z_5
p_i	0.9	0.8	0.9	0.93	0.91
$\pi_i = 1 - p_i$	0.1	0.2	0.1	0.07	0.09
π_i^2	0.01	0.04	0.01	0.0049	0.0081
$\overline{p}_i = 1 - \pi_i^2$	0.99	0.96	0.99	0.9951	0.9919
c_i	80	30	35	50	20
Q_i	1666	250	333	1613	989
$q_i = Q_i \pi_i$	166	50	33	113	89
$\overline{q}_i = Q_i \pi_i^2$	16.6	10	3.3	8	8
$\Delta q_i = q_i - \overline{q}_i$	150	40	30	105	81

Table 1 Characteristics of the static model

Reliability model which maximizes the failure-free probability is:

$$\log(r(x)) = \log(0.548402) + x_1 \log(0.99/0.9) + x_2 \log(0.96/0.8) + \\ + x_3 \log(0.99/0.9) + x_4 \log(0.9951/0.93) + x_5 \log(0.91/0.9919) \longrightarrow \max, \tag{8}$$

$$80x_1 + 30x_2 + 35x_3 + 50x_4 + 20x_5 \leq 100, \qquad x_1, x_2, x_3, x_4, x_5 \in \{0, 1\}. \tag{9}$$

By using standard software LINGO we get the optimal solution $x = (0, 1, 1, 0, 1)$ with the failure-free probability equal 0.789041, which is maximal. From the result follows that it has to double Z_2, Z_3, Z_5.

Model which minimizes the mean value of the total losses is:

$$z(x) = 451.9 - 150x_1 - 40x_2 - 30x_3 - 105x_4 - 81x_5 \longrightarrow \min, \tag{10}$$

$$80x_1 + 30x_2 + 35x_3 + 50x_4 + 20x_5 \leq 100, \qquad x_1, x_2, x_3, x_4, x_5 \in \{0, 1\}. \tag{11}$$

When we use again LINGO system, we get the optimal solution $x = (1, 0, 0, 0, 1)$ with the minimal value of losses 220.9 in the mean value. According to this solution only the parts Z_1 and Z_5 will be doubled. The differences in the solutions obtained above we can explain by great influence of the amount of the loss in the optimal solution in the second model. First solution $x = (0, 1, 1, 0, 1)$ means the most reliable system, but the loss is not minimal. Second solution $x = (1, 0, 0, 0, 1)$ gives us less reliable system, but the loss is minimal.

We can observe the values of reliability, mean loss and costs of reserves for different solutions in the Table 2. The values with a bullet are optimal for $K = 100$.

solution x	reliability	losses	doubling cost
(0,0,0,0,0)	0.578402	451.9	0
(0,1,1,0,1)	• 0.789041	300.3	85
(1,0,0,0,1)	0.657534	• 220.6	100
(1,1,1,1,1)	0.934507	45.9	215

Table 2 Static model - different solutions

3.2 Dynamic model

Let us suppose that the system's reliability should be optimized within a period $< 0, T >$ and the periods between the failures of the parts are exponentially distributed. Let the mean value of the period between two failures of the component Z_i be $1/\lambda_i$. The period of replacing the failed component Z_i by a new one has the fixed length t_i (see Figure 1).

If the Z_i is without a reserve, then the probability of the failure-free run due the part Z_i is

$$p_i = \exp(-\lambda_i T). \tag{12}$$

As the number of the failures within the period $< 0, T >$ is Poisson distributed with the mean value $\lambda_i T$, the mean value of the loss caused by the part Z_i in the case $x_i = 0$ is equal to

$$q_i = \lambda_i T Q_i, \text{ where the } Q_i \text{ is the loss caused by one failure of the part } Z_i. \tag{13}$$

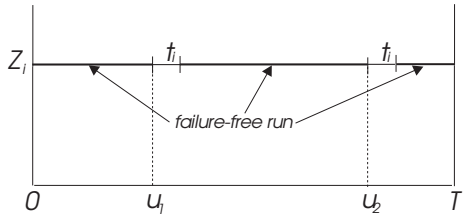


Figure 1 Two failures, u_1, u_2, in the system without a reserve part

In case that component Z_i has a reserve ($x_i = 1$), the t_i is period necessary for the reserve components' exchange, where $t_i << T$ (see Figure 2). During the period t_i the failure-free probability of the system (i.e. no failure of the reserve part) is $\exp(-\lambda_i t_i)$, see [2].

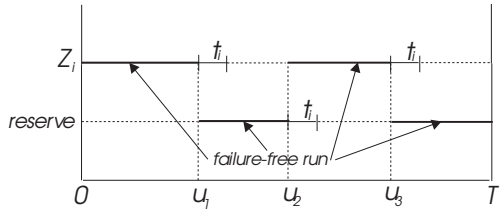


Figure 2 Failures in system with a reserve part

As the failures of a part are treated as Poisson events and therefore independent, the probability of failure-free run of the system in case of k failures of the part Z_i or its reserve is

$$(\exp(-\lambda_i t_i))^k = \exp(-k\lambda_i t_i). \tag{14}$$

Therefore, the overall probability of failure-free run of the system consisting of the part Z_i with a reserve in the period $< 0, T >$ is as follows

$$\overline{p}_i = \exp(-\lambda_i T) \sum_{k=0}^{\left[\frac{T}{t_i}\right]} \frac{(\lambda_i T)^k}{k!} \exp(-k\lambda_i t_i) = \exp(-\lambda_i T) \sum_{k=0}^{\left[\frac{T}{t_i}\right]} \frac{(\lambda_i T \cdot \exp(-\lambda_i t_i))^k}{k!} \approx \exp(\lambda_i T[\exp(-\lambda_i t_i)-1]), \tag{15}$$

where

$\left[\frac{T}{t_i}\right], \quad t_i << T$ is the top number of failures in the period $< 0, T >$,

$\exp(-\lambda_i T)\frac{(\lambda_i T)^k}{k!}$ is the probability of k failures of the part Z_i or its reserve,

Let us find the logarithm of the reliability function $r(x)$; using the general formula (5) and supposing $t_i << T$ we have

$$\log r(x) = \sum_{i=1}^{n} \log(p_i) + \sum_{i=1}^{n} x_i \log(\overline{p}_i/p_i) = \sum_{i=1}^{n}(-\lambda_i T) + \tag{16}$$

$$+ \sum_{i=1}^{n} x_i \log \sum_{k=0}^{\left[\frac{T}{t_i}\right]} \frac{(\lambda_i T)^k}{k!} \exp(-k\lambda_i t_i) \approx -T \sum_{i=1}^{n} \lambda_i + T \sum_{i=1}^{n} x_i \lambda_i \exp(-\lambda_i t_i).$$

The mean value of the number of system failures due to the component Z_i's, i.e. $\overline{q_i}$ can be established in the following way.

The number of failures both in the part Z_i and its reserve is Poisson distributed. The time between the two failures is exponentially distributed with parameter λ_i. The system fails if the time interval between the failures of the part Z_i and its reserve (or vice versa) is shorter then the repair interval t_i. The probability of this event is equal $1 - \exp(-\lambda_i t_i)$. Let the number of failures of the part Z_i in the interval $< 0, T >$ be k. These failures can be taken as binomial events, with one outcome " failure of the reserve of Z_i within interval t_i, i.e. system failure" and other outcome "no such failure within t_i. Then probability $P(X_i = r)$, where X_i is the number of system failures due to Z_i and its reserve, is as follows:

$$P(X_i = r) = \sum_{k=r}^{\left[\frac{T}{t_i}\right]} \binom{k}{r} (1 - \exp(-\lambda_i t_i))^r (\exp(-\lambda_i t_i))^{k-r} \cdot \frac{(\lambda_i T)^k}{k!} \exp(-\lambda_i T) =$$

$$= (1 - \exp(-\lambda_i t_i))^r \frac{(\lambda_i T)^r}{r!} \exp(-\lambda_i T) \sum_{k=r}^{\left[\frac{T}{t_i}\right]} \frac{(\lambda_i T)^{k-r}}{(k-r)!} \exp(-(k-r)\lambda_i t_i), \tag{17}$$

$$E(X_i) = \sum_{r=0}^{\left[\frac{T}{t_i}\right]} r \cdot P(X_i = r) = (1 - \exp(-\lambda_i t_i))\, \lambda_i T \exp(-\lambda_i T) \sum_{r=1}^{\left[\frac{T}{t_i}\right]} (1 - \exp(-\lambda_i t_i))^{r-1} \frac{(\lambda_i T)^{r-1}}{(r-1)!} \cdot$$

$$\cdot \sum_{k=r}^{\left[\frac{T}{t_i}\right]} \frac{(\lambda_i T)^{k-r}}{(k-r)!} \exp(-(k-r)\lambda_i t_i) = (1 - \exp(-\lambda_i t_i))\, \lambda_i T \exp(-\lambda_i T) \sum_{s=0}^{\left[\frac{T}{t_i}\right]-1} \frac{(\lambda_i T)^s}{s!} \approx (1 - \exp(-\lambda_i t_i))\, \lambda_i T,$$

$$\overline{q}_i = E(X_i) \cdot Q_i = (1 - \exp(-\lambda_i t_i))\, \lambda_i T Q_i. \tag{18}$$

Example 2. Let us have parts Z_1, Z_2, Z_3, Z_4, Z_5 (see Table 3 for characteristics; values of c_i and Q_i are the same as in the static example). Let the overall time be $T = 5$ [in hours]. The costs of the parts' doubling are limited by the amount $K = 100$.

Reliability model which maximizes the failure-free probability is:

$$\log(r(x)) = \sum_{i=1}^{n} \log(p_i) + \sum_{i=1}^{n} x_i \log(\overline{p}_i/p_i)] \approx -T \sum_{i=1}^{n} \lambda_i + T \sum_{i=1}^{n} x_i \lambda_i \exp(-\lambda_i t_i) = \tag{19}$$

$$= -8.5 + x_1 \cdot 1.921579 + x_2 \cdot 1.446960 + +x_3 \cdot 1.701680 + x_4 \cdot 2.160696 + x_5 \cdot 0.978240 \longrightarrow \max,$$

	Z_1	Z_2	Z_3	Z_4	Z_5
λ_i	0.4	0.3	0.35	0.45	0.2
t_i [in hours]	0.1	0.12	0.08	0.09	0.11
p_i	0.135335	0.223130	0.173774	0.105399	0.367879
$\overline{p}_i$	0.924574	0.948342	0.952828	0.914567	0.978475
$\log(\overline{p}_i/p_i)$	1.921579	1.446960	1.701680	2.160696	0.978240
c_i	80	30	35	50	20
Q_i	1666	250	333	1613	989
q_i	3332	375	582.75	3629.25	989
$\overline{q}_i$	130.649	13.2599	16.0907	144.048	21.5204
$\Delta q_i = q_i - \overline{q}_i$	3201.35	361.740	566.659	3485.20	967.480

Table 3 Characteristics of the dynamic model

$$80x_1 + 30x_2 + 35x_3 + 50x_4 + 20x_5 \leq 100, \qquad x_1, x_2, x_3, x_4, x_5 \in \{0, 1\}. \tag{20}$$

By using LINGO software we get the optimal solution $x = (0, 1, 0, 1, 1)$ with the failure-free probability equal 0,019958, which is maximal. From the result follows that it has to double Z_2, Z_4, Z_5.

Model which minimizes the mean value of the total losses is:

$$z(x) = 8908 - 3201.35x_1 - 361.740x_2 - 566.659x_3 - 3485.20x_4 - 967.480x_5 \longrightarrow \min, \tag{21}$$

$$80x_1 + 30x_2 + 35x_3 + 50x_4 + 20x_5 \leq 100, \qquad x_1, x_2, x_3, x_4, x_5 \in \{0, 1\}. \tag{22}$$

By LINGO we get we get the optimal solution $x = (0, 1, 0, 1, 1)$ with the minimum loss equal 4093.580.

In contrary to the static model, in the dynamic model the optimal solution on respect to both maximum reliability and minimum loss is the same (see Table 4 for different solutions).

solution x	reliability	losses	doubling cost
(0,0,0,0,0)	0.000203	8908	0
(0,1,0,1,1)	0.019958	4093.580	100
(1,1,1,1,1)	0.252368	325.571	215

Table 4 Dynamic model - different solutions

Acknowledgements

This research was supported by GAČR grant No. GACR P403/12/1947 and the project F4/18/2011 founded by the Internal Grant Agency of the University of Economics, Prague.

References

[1] Černý, M.: Binary Segmention and Bonferroni-Type Bounds, *Kybernetika* **47** (2011), No. 1, 38–49. ISSN 0023-5954.

[2] Dupač, V.: Private communication.

[3] Nemhauser, G. L., Wolsey, L. A.: *Integer and Combinatorical Optimization.* John Wiley & Sons Inc., New York, 1988.

Selected econometric methods of optimization of economic policy

Pavlína Pellešová[1]

Abstract. This paper deals with approaches to the optimization of economic policy using econometric methods. In particular, the econometric model is used for the selection and optimization of tools and appropriate levels of management within the framework of economic policy. Four basic types of variables used in econometric approaches to optimization of economic policy are described. Then it presents methods for the selection of optimal control and optimization of economic policy instruments, namely the method of target variables, the principles of optimal control and simulation procedures. The article is focused on the method of target variables. The selected method, its advantages and disadvantages and its possible applications for making optimal economic policy are dealt with.

Keywords: econometric methods, optimization of economic policy, method of target variables;

JEL Classification: C49

AMS Classification: 91G70, 91B82

1 Econometric methods of optimization of economic policy

Econometrics is based upon the development of statistical methods for estimating economic relationships, testing economic theories, and evaluating and implementing government and business policy. The most common application of econometrics is the forecasting of such important macroeconomic variables as interest rate, inflation rate and gross domestic product.[2] It is possible to use econometric methods, econometric modelling, especially for the selection and optimization of tools and appropriate levels of management for optimization of the economic policy. In the area of economic optimization an econometric model is the solution of the problem at the corporate level and also at the macroeconomic level. The essence of the model is to predict the influence or the impact of possible options of economic policy on the functioning of the economic system under examination. If criterion for assessing the relative advantage of alternative management tools is known, it is possible to choose from a set of solutions the right one to ensure the achievement of the goal, while respecting all constraints.

1.1 Variables in econometric approaches of optimization

Objectives of short-term economic policy are realized by stabilization measures, but long-term goals are realized on the basis of strategic decisions based on pre-approved program. The choice of instruments of economic policy and their optimization are closely linked to economic forecasting in the form of feedback. Conditional forecast is based inter alia on expected or proposed decision of management entity. In determining the specific economic policy the anticipated impact of the expected economic decisions or actions must also be taken into account.

Short-term economic policy uses macroeconomic model to select the instruments of economic regulation to ensure stability or required economic development generally for a period of one year. Long-term economic policy aims to ensure economic growth in a period of five years or more. Therefore, in optimization of econometric approaches to economic policy there are four basic types of variables:

- target variables - are controlled by endogenous variables, quantify the targets, other endogenous variables are in determining economic policy irrelevant;
- control variables - exogenous variables whose values are affected by managing entity, i.e. economic policy; autonomous variables i.e. remaining exogenous and other lagged exogenous variables are not under the control of the managing entity and have the form of input data;
- target endogenous variables are associated with the main objectives of modelled economic system, they are the mostly macroeconomic variables (GDP, aggregate consumption, investment, unemployment, inflation, trade balance and balance of payments or growth of these indicators);

[1] Silesian University in Opava, School of Business Administration in Karvina, Department of Economy, Univerzitní nám. 1934/3, 733 40 Karviná, Czech Republic

[2] WOOLDRIDGE, J. M. *Introductory Econometrics – A modern approach.* 4th edition. USA: South-Western Cengage Learning, 2009, p. 1. ISBN-13 978-0324581621.

- control exogenous variables - used to realization of the intentions of macroeconomic regulation, they are the instruments of fiscal policy, monetary policy tools, instruments for external economic policy.

The econometric approach to variant setting of the economic policy or management strategy is based on connecting the information, obtained from estimated econometric model, with additional information regarding the object of regulation or control. The ways of the selection from a number of possible variants of economic policy with regard to the chosen objective criterion are several.

1.2 Dynamic model of simultaneous equations

For selection and evaluation of different options of economic policy based on a dynamic model of simultaneous equations (MSR) in structural form for any period t (where y_t is G-dimensional vector of target endogenous variables, y_{t-1} is G-dimensional vector of lagged target endogenous variables, x_{t-1} is m-dimensional vector of lagged control exogenous variables, z_t is K-dimensional vector of autonomous exogenous variables, u_t is G-dimensional vector of random components of the usual properties; $B, \Gamma_1, \Gamma_2, \Gamma_3$ - are matrices of structural parameters):

$$B \cdot y_t + \Gamma_1 \cdot y_{t-1} + \Gamma_2 \cdot x_{t-1} + \Gamma_3 \cdot z_t = u_t, \qquad t = 1,2,...,T. \tag{1}$$

For a regular matrix B is possible to express a vector of target variables y_t in the reduced form

$$y_t = \Pi_1 \cdot y_{t-1} + \Pi_2 \cdot x_{t-1} + \Pi_3 \cdot z_t + v_t, \qquad t = 1,2,...,T, \tag{2}$$

where $v_t = B^{-1} \cdot u_t$ is G-dimensional vector of random components of the usual properties, $\Pi_1 = -B^{-1} \cdot \Gamma_1$ is matrix of dynamic multipliers of size (G,G), $\Pi_2 = -B^{-1} \cdot \Gamma_2$ is matrix of dynamic multipliers of size (G,m), and $\Pi_3 = -B^{-1} \cdot \Gamma_3$ is matrix of common multipliers of size (G,K).

For the selection and optimization of economic policy instruments are often used:

- method of target variables;
- principles of optimal control;
- simulation procedures.

The method of target variables will be described in detail. All these methods for determining the optimal economic policy assume that decision makers know the consistently estimated values of the parameters economic model. For determining the optimal economic policy ensuring the achievement of these goals, it is possible to proceed in two possible ways. First, the so-called open procedure consists in determining the time sequence or trajectory of the optimal values of control exogenous variables at the beginning of time of management, for example, the length h, i.e. regardless the future development of target variables. The result is a sequence of vectors of the optimum values of economic management tools (which guarantees the attainment of desired extreme of selected criteria functions):

$$x_{T+1}^0, x_{T+2}^0, ..., x_{T+h}^0 \tag{3}$$

The control with feedback is the second method of optimal control, based on the assumption that option of economic policy chosen in any period affects management strategy in future periods. The proposed level and structure of economic policy instruments are influenced by the impact of specific economic policy on the controlled target variables. If we respect the link between management tools used and results achieved, in any period T we can express vector of the optimal values of control variables x_T^0 as a function of the actually achieved values of target variables in the previous period $T-1$, so we use the information about detected outcomes of procedure in each period to correcting of calculation of the optimal strategy for the next period.

2 Method of target variables

Once econometric models in the form of a system of simultaneous equations were used to study likely consequences of alternative policies, it was a natural step to make such studies more systematic. Besides the econometric model one would specify an objective function involving both endogenous and exogenous variables.[3] The approach based on the using of control and target variables of the simultaneous equations in setting economic strategy is known as Tinbergen's procedure. This procedure is based on, in determining the optimal values of the instruments of economic policy, the assumption that critical entity knows constant desired level of selected target endogenous variables for each period. The vector in next period $T+1$ marks y_{T+1}^{**} , the number of control exogenous variables is at least equal to the number of target variables, or if $m \geq G.$ An estimated structural shape of model of simultaneous equations with using known goals of economic policy in the period $T+1$ can be written as:

$$\hat{B}y_{T+1}^{**} + \hat{\Gamma}_1 \cdot y_T + \hat{\Gamma}_2 \cdot x_T + \hat{\Gamma}_3 \cdot \hat{z}_{T+1} = e_{T+1}, \tag{4}$$

where $\hat{B}, \hat{\Gamma}_1, \hat{\Gamma}_2, \hat{\Gamma}_3$ are consistent estimates of matrices of structural parameters, $\hat{z}_{T+1}$ - is a vector of estimates of autonomous exogenous variables, e_{T+1} is a vector of residuals. Instead of y_T we can also write y_T^{**} , because only in initial period of horizon $h,$ is the vector of true values, while in all other periods there are required values for the previous period. If the requirement $m = G$ is fulfilled or the so-called number of degrees of freedom of economic policy $m - G$ is zero, during which time the matrix Γ_2 is regular, we obtain the solution of the previous equation according to vector of control variables with clearly the optimal values of the individual instruments of economic policy x_T^0 in the form:

$$x_T^0 = -\Gamma_2^{-1} \cdot B \cdot y_{T+1}^{**} - \Gamma_2^{-1} \cdot \Gamma_1 \cdot y_T - \Gamma_2^{-1} \cdot \Gamma_3 \cdot z_{T+1} + \Gamma_2^{-1} \cdot e_{T+1}, \tag{5}$$

In this equation, the target endogenous variables and exogenous control variables exchanged roles, so the optimal values of economic instruments are expressed as a linear function of the desired values of target variables in period $t+1$ and values of residues in the same period. From the equation is clear mutual dependence of economic instruments and desired goals, because in general, the optimal level of each of the control variables depends on the values of all target variables. Only in special cases the specific control variable has influence only on one or more of the total number of target endogenous variables. If it is $m > G$, there are multiple solutions according to the vector of control variables, but it is possible to proceed so that for $m - G$ we put a pre-agreed fixed values and solve set for the remaining G of unknown instruments of economic policy. For $m < G$ there is no solution. Using the equation we can examine the sensitivity of the optimal values of individual control variables to changes of each from the explanatory variables of this equation. For example the marginal values

$$\frac{\partial x_T^0}{\partial y_{T+1}^{**}} = -\Gamma_2^{-1} \cdot B, \tag{6}$$

express reaction of the optimal management tools in the period T to changes in the desired values of target variables for the period $T+1$. It is an analogy of multiplier, which indicates the average reaction of explained endogenous variables to unit change of some from the control exogenous variables, et ceteris paribus.

As an illustration of described one-off procedure we use the simple Keynesian macroeconomic model (where C - final consumption, Y - income, Z - autonomous investment and public expenditure, u - random component of the usual properties):

$$C_t = \beta_1 + \beta_2 Y_t + \beta_3 Y_{t-1} e_{T+1} + u_t, \tag{7}$$

$$Y_t = C_t + Z_t, \tag{8}$$

[3] CHOW, C. G. Econometrics and economic policy. *Statistica Sinica 11*. Taiwan: Academia Sinica, 2001, p. 638. ISSN 1017-0405.

The adjusted (reduced) form of equation for income can be written as follows:

$$Y_t = \pi_1 + \pi_2 Y_{t-1} + \pi_3 Z_t + v_t, \tag{9}$$

If we know the consistent estimates $\hat{\pi}_1, \hat{\pi}_2, \hat{\pi}_3$, obtained from T observations, we may express income for the period $T+1$ as follows:

$$Y_{t+1} = \hat{\pi}_1 + \hat{\pi}_2 Y_T + \hat{\pi}_3 Z_T + \hat{v}_{T+1} \tag{10}$$

Assume that $Z_{T+1} = Z_T$. For a given desired value of the target endogenous variables Y_{+1}^{**} we determine the corresponding optimal value of the control variable Z_T^0, which is a tool of the short-term economic policy and the solution of the relationship according to Z_T $(m = G = 1)$ is:

$$Z_T^0 = \frac{Y_{T+1}^{**} - \pi_1 - \pi_2 \cdot Y_T - v_{T+1}}{\pi_3}. \tag{11}$$

The expected reaction of optimal level of autonomous investment and public expenditures to unit change of the desired target value of income is given by expression, which corresponds to the reciprocal value of the multiplier of income considering exogenous variable Z :

$$\frac{\partial Z_T^0}{\partial Y_{T+1}^{**}} = \pi_3^{-1}. \tag{12}$$

If we express without time resolution reduced form of vector of target variables after the merger of the vector of target lagged endogenous variables and vector of autonomous exogenous variables into one vector of predetermined variables, e. g. r , we get the shape, where $\hat{\Pi}_*$ and $\hat{\Pi}_{**}$ are matrices of size (G, m), respectively $(G, G+K)$ of known or estimated multiplier:

$$y = \hat{\Pi}_* x + \hat{\Pi}_{**} r + \hat{v}. \tag{13}$$

To determine vector x^0 of the optimal values of the instruments of economic policy for one arbitrary period, which guarantee the achievement of the desired values of vector of target variables y^{**} without respect of the random effects, we base on deterministic version of the previous relationship. After the substitution $q = \hat{\Pi}_{**} r$ it can be rewritten in the form:

$$y = \hat{\Pi}_{**} x + q \tag{14}$$

If the matrix of multiplier $\hat{\Pi}_*$ square $(m = G)$ and regular, after substituting y^{**} and x^0 for y , respectively x , we determine single optimal economic policies providing the required targets with solution transcribed relationship that has the form:

$$x^0 = \hat{\Pi}_*^{-1}(y^{**} - q). \tag{15}$$

The procedure and interpretation of solutions of this relationship in special cases, where matrix $\hat{\Pi}_*$ is diagonal or triangular, was stated for example by Friedman. Diagonal matrix corresponds to a situation in which each control variable has influence to only one target variable, where $m = G$. In the case of a triangular matrix $\hat{\Pi}_*$ the control variable x_i has influence on only one target variable y_i, other control variable x_j affects the target variables y_i and y_j. However, matrix defined in this way is not usually real in practice. Presented procedure at one-off determining of optimum values of economic policy instruments can be generalized for the dynamic version of the researched task, when there is determination of long-term economic policy for several seasons in advance.

3 Conclusion

The disadvantage of Tinbergen's procedure for finding the optimal management strategy using method of target variables is mainly the fact that it does not allow possibility of interaction and compensation of changes in the values of different objectives, but for each of them requires a constant value. Restrictive is the condition of existence of a sufficient number of economic management tools $m \geq G$, as well as the exact identification of all target values of control endogenous variables in advance, and leaving to the discretion the determining the length of horizon of control.

If we compare Theil's specification of loss, respectively preference function, it is based on the assumption that target endogenous variables affect the level directly, while the exogenous variables can affect it only indirectly. Compared to the method of target variables the optimization procedure is not bound to comply with the requirements $m \geq G$ and allows interaction between levels of different target variables. This means that the decline in the value of one target variable can compensate for the growth of another, without changing the level of preference functions. Not only this procedure but also other methods of optimal control in econometrics have against the Tinbergen's procedure advantage in that it is not necessary to determine the exact values of target variables in advance in any period. The advantage of optimal control methods is the fact that the factor of uncertainty, represented by random components of the model of simultaneous equations can be easily included in the optimal solution.

References

[1] Chow, C. G.: *Econometrics and economic policy. Statistica Sinica 11*. Taiwan: Academia Sinica, 2001.
[2] Klein, M. W.: *Mathematical methods for economics.* McGraw-Hill, 2009.
[3] Marek, L. a kol.: *Statistika pro ekonomy – aplikace.* Professional Publishing, Praha, 2007.
[4] Mielcova, E.: *Teorie her a ekonomického rozhodování*. Karviná: OPF SU, 2011.
[5] Paulík, T.: *Hospodářská politika*. Frýdek-Místek: Kleinwachter, 2002.
[6] Ramík, J.: *Ekonomicko-matematické metody*. Karviná, OPF SU, 2008.
[7] Vincúr, P., Zajac, Š.: *Úvod do prognostiky.* SPRINT, Bratislava, 2009.
[8] Wooldridge, J.M.: *Introductory Econometrics - A Modern Approach.* 4. Edition. USA: South-Western Cengage Learning, 2009.

DAME - Microsoft Excel add-in for solving multicriteria decision problems with scenarios

Radomir Perzina[1], Jaroslav Ramik[2]

Abstract. The main goal of every economic agent is to make a good decision, especially in economic environment with many investment alternatives and evaluation criteria. The Analytic Hierarchy Process (AHP) is widely used approach for solving decision making problems. There exists wide range of computer programs that are able to help decision makers to make good decisions. 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.
This paper introduces 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, 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 – Saaty's Method, Geometric Mean Method and Fuller's Triangle Method. All calculations are instant so users can easily see what happen if value of any input is changed. Apart from the final ordering of the variants there are also shown all intermediate results, so it is clearly seen how the synthesis was produced. The results of the decision model are depicted by a bar chart. Capabilities of the proposed software package are demonstrated on couple of illustrating examples of real life decision problems.

Keywords: analytic hierarchy process, multi-criteria decision making, pair-wise comparisons, Microsoft Excel, Scenarios.

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]. 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, 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 - Saaty's Method [10], Geometric Mean Method [1] and Fuller's Triangle Method [2].

[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.

2 Software description

DAME works with all current versions of Microsoft Excel from version 97. It consists of four individual files:

- DAME.xla – main module with user interface, it is written in VBA (Visual Basic for Applications),
- DAME.dll – it contains special functions used by the application, it is written in C#,
- DAME.xll – it contains library for linking C# modules with Excel called Excel-DNA (http://exceldna.codeplex.com),
- DAME.dna – configuration file for Excel-DNA module.

All four files must be placed in the same folder and macros must be permitted before running the module (see 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.

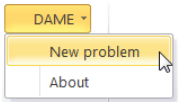


Figure 1 New problem menu

Then there will be shown a form with main problem characteristics, see figure 2.

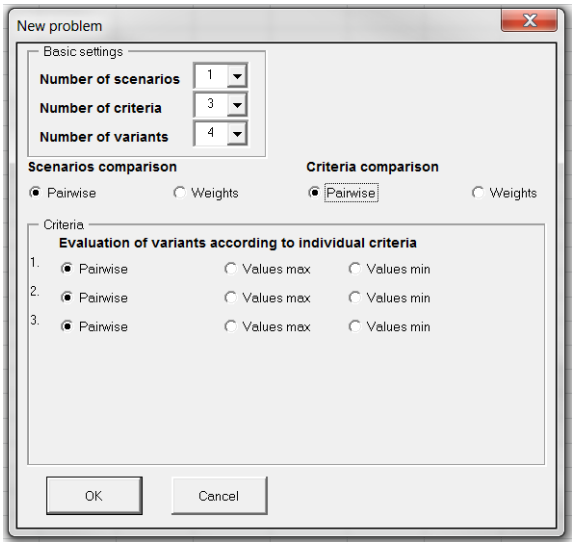


Figure 2 New problem characteristics

In the top panel there are basic settings: Number of scenarios, criteria and variants. In case a user doesn't want to use scenarios, the number of scenarios should be set to one. In the second panel we can set how we want to compare scenarios and criteria either using pairwise comparison matrix or set weights directly. 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 criteria (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" car with 3 decision criteria and 3 variants. The goal of this realistic decision situation is to find the best variant from 3 preselected ones according to 3 criteria: price (minimization criterion), efficiency (pairwise) and design (pairwise). At this stage we don't use scenarios, so number of scenarios will be set to one. Setting of parameters can be seen on the figure 4.

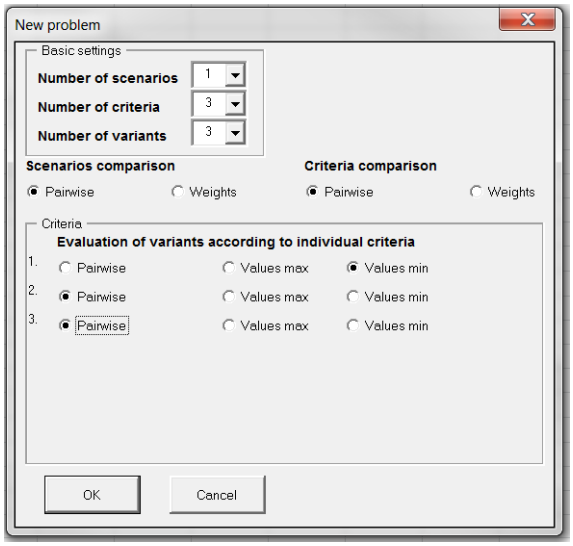


Figure 4 Case study – setting of parameters

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	0.005	*Criteria weights*
price	1	2	3		0.539615
efficiency	0.5	1	2		0.296961
design	0.333333	0.5	1		0.163424

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	260				0.292405	0.625013	0.167656
Var 2	220				0.34557	0.238487	0.483604
Var 3	210				0.362025	0.1365	0.348739

efficiency	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	

design	Var 1	Var 2	Var 3	0.076
Var 1	1	1/2	1/3	
Var 2	2	1	2	
Var 3	3	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 criteria (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.

CZn=	Weight	Rank
Var 1	0.37079	1
Var 2	0.336328	2
Var 3	0.292882	3

Figure 8 Case study – total evaluation of variants

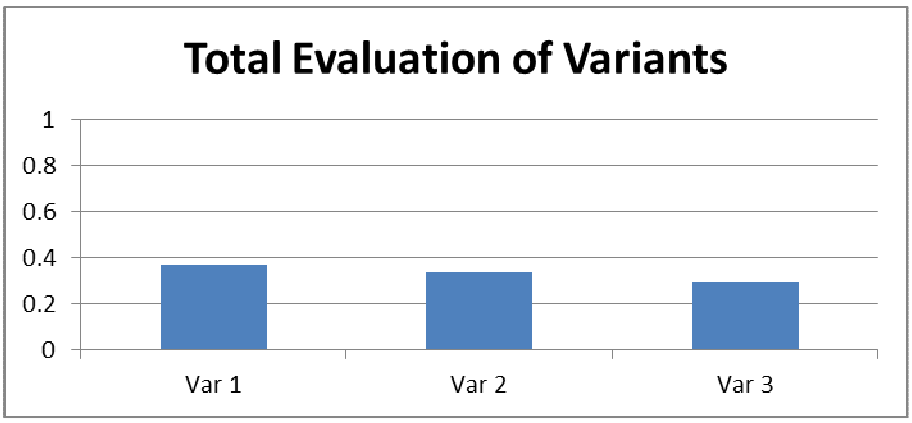


Figure 9 Case study – total evaluation of variants - graph

4 Case Study with Scenarios

In real decision situations a decision maker usually faces uncertainty. For example it may happen that price goes up or efficiency is calculated based on special conditions that are far from real ones. That is why our proposed software works also with scenarios. In this case study we assume two scenarios – optimistic and pessimistic. First we must compare both scenarios using pairwise comparison matrix. It can be seen on figure 10.

Scenarios	optimistic	pessimistic	0.000	*Scenarios weights*
optimistic	1	2		0.666667
pessimistic	0.5	1		0.333333

Figure 10 Case study – scenarios comparison

Optimistic scenario is using exactly the same entries as in the previous case study, so we need to just evaluate variants to individual criteria for the second - pessimistic scenario, see figure 11.

price	Value				*Variants weights*		
Var 1	320				0.276753	0.412599	0.124306
Var 2	240				0.369004	0.32748	0.517134
Var 3	250				0.354244	0.259921	0.35856

efficiency	Var 1	Var 2	Var 3	0.030
Var 1	1	1	2	
Var 2	1	1	1	
Var 3	0.5	1	1	

design	Var 1	Var 2	Var 3	0.060
Var 1	1	1/3	1/4	
Var 2	3	1	2	
Var 3	4	0.5	1	

Figure 11 Case study – evaluation of variants – pessimistic scenario

Final evaluation of variants for pessimistic scenario can be seen on figure 12.

Zn=	Weight	Rank
Var 1	0.2921803	3
Var 2	0.3808808	1
Var 3	0.3269389	2

Figure 12 Case study – final evaluation of variants – pessimistic scenario

Finally from both scenarios there is calculated synthesis and total evaluation of variants is shown on figure 13.

CZn=	Weight	Rank
Var 1	0.3445867	2
Var 2	0.3511792	1
Var 3	0.3042341	3

Figure 13 Case study – total evaluation of variants

Comparing to the previous case study without scenarios we can see that final rank of variants was changed. Now the best variant is Var 2 with weight 0.35, then Var 1 with weight 0.34 and the last one Var 3 with weight 0.30.

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. The new add-in could be used mainly by students, scientists and small companies.

Acknowledgements

This research was supported by the grant project of GACR No. 402090405.

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] Ramík, J., Vlach, M.: *Generalized concavity in optimization and decision making*. Kluwer Publ. Comp., Boston-Dordrecht-London, 2001, 305 p.
[9] Saaty, T.L.: Exploring the interface between hierarchies, multiple objectives and fuzzy sets. *Fuzzy Sets and Systems* 1, 1978, 57-68.
[10] Saaty, T.L.: *Multicriteria decision making - the Analytical Hierarchy Process*. Vol. I., RWS Publications, Pittsburgh, 1991.
[11] Saaty, T.L.: *Decision Making with Dependence and Feedback – The Analytic Network Process*. RWS Publications, Pittsburgh, 2001.

Max-plus algebra at road transportation

Štefan Peško[1], Michal Turek[2], Richard Turek[3]

Abstract. In this paper two models of two transportation problems over the max-plus algebra are analysed. First problem deals with coordination of strongly dependent light crossroads. Second problem focuses on computing bus line timetables and on synchronization of departures from some interchange stops at bus transportation network. The behaviour of this discrete-event dynamic systems can be modeled as linear systems in max-plus algebra with operations max and plus. We show how it is possible to use eigenvalues and eigenvectors of matrix for computing practical characteristics of these road transport systems. Some results of computation experiments using open source software ScicosLab with real data of Czech town Protějov instances of problems are presented.

Keywords: max-plus algebra, eigenproblems, discrete-event dynamic systems, light crossroads, bus line timetables

JEL classification: C51, R48
AMS classification: 15A06, 90C48,

1 Introduction

Max-plus algebra is an attractive way to describe a class of non-linear problems that appear for instance in discrete event dynamic systems [11, 5]. In this paper two models of two transportation problems over max-plus algebra are analysed. First problem deals with coordination of strongly dependent light crossroads [7]. Second problem focuses on computing bus line timetables and on synchronization of departures from some interchange stops at bus transportation network [9]. Similar approach based on eigenproblem can be found in [3]. We show how it is possible to use eigenvalues and eigenvectors of matrix for optimizing practical characteristics of these road transport systems. Every computation is done via open software ScicosLab [6]. We begin with some known facts from the theory of the max-plus algebra.

2 Max-plus algebra

We will use needed notation and formulation from [1, 4]. We suppose that the reader is not familiar with the basic definition for max-plus algebra. An extensive discution of the max-plus algebra can be found in [4].

Let $\Re$ be the set of real numbers, $\varepsilon = -\infty, e = 0$, $\Re_{\max} = \Re \cup \{\varepsilon\}$, $\underline{n} = \{1, 2, \ldots, n\}$ where $n \geq 2$ is a natural number.

Let $a, b \in \Re_{\max}$ and define operations $\oplus$ and $\otimes$ by: $a \oplus b = \max(a, b)$ and $a \otimes b = a + b$. Let $\Re_{\max}^{n\times n}$ be set of $n \times n$ matrices with coeficients in $\Re_{\max}$. Matrices can be added and multiplied formally in the same manner as in the classical algebra. The sum of matrices $\mathbf{A}, \mathbf{B} \in \Re_{\max}^{n\times n}$ denoted by $\mathbf{A} \oplus \mathbf{B}$ is defined by $(\mathbf{A} \oplus \mathbf{B})_{ij} = a_{ij} \oplus b_{ij} = \max\{a_{ij}, b_{ij}\}$ for $i, j \in \underline{n}$. The product of matrices $\mathbf{A} \in \Re_{\max}^{n\times l}, \mathbf{B} \in \Re_{\max}^{l\times n}$ denoted by $\mathbf{A} \otimes \mathbf{B}$ is defined by $(\mathbf{A} \otimes \mathbf{B})_{ij} = \bigoplus_{k=1}^{l} a_{ik} \oplus b_{kj}$ for $i, j \in \underline{n}$.

The communication graph of matrix plays in our models the fundamental role.

[1]University of Žilina, Faculty of Management Science and Informatics, Univerzitná 8215/1, 010 26 Žilina, Slovakia, e-mail: stefan.pesko@fri.uniza.sk

[2]VŠB - Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, 70833 Ostrava-Poruba, Czech Republic, e-mail: michal.turek.st1@vsb.cz

[3]VŠB - Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, 70833 Ostrava-Poruba, Czech Republic, e-mail: richard.turek.st@vsb.cz

Let $\mathbf{A} \in \Re_{\max}^{n\times n}$ be any matrix. A weighted digraph $\mathcal{G}(\mathbf{A}) = \big(\mathcal{N}(\mathbf{A}, \mathcal{D}(\mathbf{A})\big)$, where $\mathcal{N}(\mathbf{A}) = \underline{n}$ is set of nodes, $\mathcal{D}(\mathbf{A}) = \{(i,j) \in \underline{n} \times \underline{n} : a_{ji} \neq \varepsilon\}$ is set of arcs and a weight $w(i,j) = a_{ji}$ is associated with any arc is called **communication graph** of $\mathbf{A}$.

A path from node i to node j is a sequence of arcs $p = \{(i_k, j_k) \in \mathcal{D}(\mathbf{A})\}_{k\in\underline{m}}$ such that $i = i_1, j_k = i_{k+1}$ for $k < m$ and $j_m = j$. The path p consists of the nodes $i = i_1, i_2 \ldots, i_m, j_m = j$ with length m denoted by $|p|_1 = m$. In the case when $i = j$ the path is said to be circuit. A circuit is said to be cycle if nodes i_k and i_l are different for $k \neq l$.

Let us denote by $P(i,j;m)$ the set of all paths from node i to node j of length $m \geq 1$ and for any arcs $(i,j) \in \mathcal{D}(\mathbf{A})$ let its weight be given by a_{ji}. Then weight of path $p \in P(i,j;m)$ denoted by $|p|_w$ is defined to be the sum of the weights of all the arcs that belong to the path. The average weight of path p is given by $|p|_w/|p|_1$.

Let $\mathbf{A} \in \Re_{\max}^{n\times n}$ be a matrix. If $\mu \in \Re_{\max}$ is a scalar and $\mathbf{v} \in \Re_{\max}^{n}$ is a vector of at least one finite element such that: $\mathbf{A} \otimes \mathbf{v} = \mu \otimes \mathbf{v}$ then, μ is called an **eigenvalue** and $\mathbf{v}$ an **eigenvector**.

Let $C(\mathbf{A})$ denote the set of all cycles in $\mathcal{G}(\mathbf{A})$ and write: $\lambda(\mathbf{A}) = \max_{p\in C(\mathbf{A})} \frac{|p|_w}{|p|_1}$ for the maximal everage cycle weight. An efficient way of evolution $\lambda(\mathbf{A})$ is Karp's algorithm of complexity $O(n^3)$ or numericaly almost linear Howard's algorithm. A cycle $p \in \mathcal{G}(\mathbf{A})$ is said to be **critical** if its average weight is maximal.

Theorem 1 (Bacceli at al. [1]). *Let $\mathbf{A} \in \Re_{\max}^{n\times n}$ is irreducible or equivalent if $\mathcal{D}(\mathbf{A})$ is strongly conected. Then one and only one finite eigenvalue (with possible several eigenvectors) exists. This eigenvalue is equal to the maximal average weight of cycles in $\mathcal{G}(\mathbf{A})$.*

Cuninghame-Green (1960) showed (cited in [2]) that $\lambda(\mathbf{A})$ is an optimal solution of the linear program

$$\begin{aligned} \lambda &\rightarrow \min && (1)\\ \lambda + x_i - x_j &\geq a_{ij} \qquad \forall (i,j) \in \mathcal{D}(\mathbf{A}), && (2)\\ x_i &\geq 0 \qquad \forall i \in \mathcal{N}(\mathbf{A}). && (3) \end{aligned}$$

Because $\lambda(\mathbf{A})$ is the optimal value of linear program (1-3), following fact is proven.

Theorem 2 (Cechlárová [2]). *Let $\mathbf{A} \in \Re_{\max}^{n\times n}$ be a matrix. Then inequality*

$$\mathbf{A} \otimes \mathbf{x} \leq \mu \otimes \mathbf{x} \tag{4}$$

is solvable if and only if $\mu \geq \lambda(\mathbf{A})$.

Following max-plus models use two practical interpretation of the eigenvalue and eigenvector of matrix.

- If the road transport system is performed in cycles and consists of jobs with matrix desribing duration of operations $\mathbf{A}$ that were started according to some eigenvector $\mathbf{x}$ of this matrix, then it will move forward in regular steps. The time elapsed between the consecutive starts of all jobs will be equal to $\lambda(\mathbf{A})$.
- Let us have a schedule for the road transport system that requires time interval between two consecutive jobs smaller than certain value μ. In this case idle times of system can be used to optimize some characteristic of transport elements.

  Let $\mathbf{A} \in \Re_{\max}^{n\times n}, \mu \in \Re$ from Theorem 2 and $f : \Re_{\max}^{n} \rightarrow \Re$ be a real max-plus function. Then we woud like to solve following optimization problem:

$$\begin{aligned} f(\mathbf{x}) &\rightarrow \min && (5)\\ \mathbf{A} \otimes \mathbf{x} &\leq \mu \otimes \mathbf{x}. && (6) \end{aligned}$$

Remark 1. We don't know any publications in this area for max-plus algebra solving problems like (5-6). First computational experiments show that the critical cycle of the communication graph of $\mathbf{A}$ can play fundamental role.

3 Coordination of strongly dependent light crossroads

Managing traffic flow in congested networks requires a coordination of light crossroads and so controlling vehicular and walkers behaviour. In this section we present a basic max-plus model for the coordination of q dependent light crossroads where the constant output transport flow from one model is the input transport flow to the other and vice versa.

We focus on following formulation of problem: *Let's have scheme for strongly dependent light crossroads with traffic flows of vehicles and flows of walkers on crosswalks. Let a phase schemas for first and second crossroad be given. Our goal is to find a traffic light policy for crossings - signal schedule with minimum length of cycle and coordened relations between chosen vehicle flows.*

Inputs for the following mathematical model are:

- q – number of crossroads,
- n – number of transport flows,
- m – number of phases of the signal schedule,
- $\mathcal{P}$ – set of transport flows (vehicles and walkers),
- P_i – i^{th} transport flow, $P_i \in \mathcal{P}, i \in \underline{n}$,
- $\mathcal{F}$ – set of phases of the signal schedule,
- F_j^k – j^{th} phase of the signal schedule for k^{th} crossroad, $F_j \in \mathcal{F}, j \in \underline{m}$,
- $m(P_iP_j)$ – mean time between lock-up of P_i flow and lead-off P_j flow,
- $t(P_i)$ – minimum green time for flow P_i,
- τ – minimum green time for coordened flow.

Variables for this model are:

- λ_S – length of the cycle of the signal schedule,
- $x_i(k)$ – begining time of green for the flow P_i in k^{th} phase of the signal schedule.

Let $S_1 = \{(P_i, P_j) \in \mathcal{P} \times \mathcal{P} : P_i \in F_r^c, P_j \in F^c_{(r \mod n)+1}, c \in \underline{2}, r \in \underline{q}\}$ be the set of pair of flows between adjacent phases and let subset S_2 of the coordinated pair of flows be defined similarly. Now we can define matrix $\mathbf{A} \in \Re_{\max}^{n\times n}$ where finite elements a_{ij} give the time between lead-off transport flow P_j and lock-up transport flow P_i:

$$a_{ij} = \begin{cases} m(P_j, P_i) + t(P_j) & \text{if } (P_j, P_i) \in S_1 \\ m(P_j, P_i) + \tau & \text{if } (P_j, P_i) \in S_2 \\ \varepsilon & \text{otherwise} \end{cases} \tag{7}$$

Details can be found in [8] where author studies three types of coordened vehicle flows for two crossings scheme with two crossroads.

Solving equation of the spectral problem $\mathbf{A} \otimes \mathbf{v} = \mu \otimes \mathbf{v}$ we can set $\lambda_S = \mu$ and $\mathbf{x}(0) = \mathbf{v}$. Then the development of the system described above can be expressed by a vector equation of the form

$$\mathbf{x}(k+1) = \mathbf{A} \otimes \mathbf{x}(k).$$

When we set $\mu > \lambda(\mathbf{A})$ then we can use difference $\mu - \lambda(\mathbf{A})$ of coordined flows of crossroads as removind difference between starting green as much as possible while solving optimization problem:

$$\sum_{(P_i,P_j)\in S_2} (x_j - x_i)^2 \quad \to \quad \min \tag{8}$$

$$\mathbf{A} \otimes \mathbf{x} \quad \leq \quad \mu \otimes \mathbf{x}. \tag{9}$$

Remark 2. In general, for solving these problems conditions described in **Remark 1** also apply. In case of searching for one non-negative solution it is possible to formulate this as a problem of quadratic programming.

The earliest computational experiments were done on real data of two dependent crossroads with 12 vehicles flows and 4 walker flows in this way: First each of feasible scheme of phases for single crossroads is generated and each of feasible matrices $\mathbf{A}$ is examined. After that corresponding eigenproblems are solved. Eigenvectors are compared by objective function (8). Via perturbation scheme based on penalize of finite entry a_{ij} defined as $a_{ij} + \delta$ with randomly penalization value δ defined as $0 < \delta \leq \mu - \lambda(\mathbf{A})$ relevant eigenproblem is solved. The best of these solutions is an approximate solution of the problem (8–9). Next computational experiments were done for 3 real types of coordened flows by presented heuristic method. Experiments show that proposed difference $\mu - \lambda(\mathbf{A})$ for our instances can be used so that $f(\mathbf{x})$ object function of (8) can sometimes decrement even to the value $\frac{\lambda(\mathbf{A})f(\mathbf{v})}{\mu}$, where $\mathbf{v}$ denotes the best found eigenvector.

4 Synchronization of departures at bus line timetables

This model focuses on computing bus line timetables and on sychronisation of departures from some crossing stops at transportation network. *Let bus network with the set of bus lines, the sets of stops and crossing stops of bus lines be given. We know traveling time on line directions and turnaround time of the lines. We suppose that every line is served by given number of buses and buses do not change lines. The goal is to find the synchronized departure time of timetables when passenger changes lines on crossing stops. The objective function, the maximal difference between departure times of the same directions, is minimized.*

Inputs for the following mathematical model are:

- S – set of stops of bus network,
- S_c – set of crossing stops of bus network; $\emptyset \neq S_c \subset S$,
- q – number of lines in the bus network,
- n – number of line directions in the bus network,
- i – index of line direction; $i \in \underline{n}$,
- α_i – first stop of the line direction i; $\alpha_i \in S$,
- β_i – last stop of the line direction i; $\beta_i \in S$,
- t_i – traveling time on direction i from α_i to β_i; $i \in \underline{n}$,
- L_q – set of directions for q^{th} line; $q \in \underline{q}$,
- o_q – turnaround time of the q^{th} line,
- m_q – number of buses on the q^{th} line,

and variables are:

- $x_i(k)$ – k^{th} bus synchronized departure time in direction i of timetable,
- λ_D – length of the mean period between departure times of timetable.

Now we can define matrix $\mathbf{A} \in \Re_{\max}^{n \times n}$ where finite elements a_{ij} give the time between departure time from stop a_j and direction j and arrival time to stop β_i of direction i.

$$a_{ij} = \begin{cases} \min\{t_j, o_q/m_q\} & \text{if } \beta_i = \alpha_j, i \in L_q, j \in L_q \\ t_j/m_q & \text{if } \beta_i = \alpha_j, i \in L_q, j \notin L_q \\ o_q & \text{if } i = j, i \in L_q, \alpha_i \in S - S_c \\ \varepsilon & \text{otherwise} \end{cases} \tag{10}$$

More details can reader find in Turek's disertation work [10]. Because vector $\mathbf{x}(k) = \big(x_1(k), x_2(k), \ldots, x_n(k)\big)$ denotes the time in which all buses started for the k^{th} cycle then if all buses wait for all preceding jobs to finish their operations, the earliest possible starting time at $(k+1)^{th}$ cycle are expressed by vector $\mathbf{x}(k+1)$, which can be expressed by vector equation ower max-plus algebra in the form

$$\mathbf{x}(k+1) = \mathbf{A} \otimes \mathbf{x}(k). \tag{11}$$

If an initial condition

$$\mathbf{x}(0) = \mathbf{x}, \tag{12}$$

is given, the whole future evolution of (11) is determined. When $\mathbf{x}$ is an eigenvector and λ_D is an eigenvalue of the matrix $\mathbf{A}$ then in general

$$\mathbf{x}(k+1) = \mathbf{A}^k \otimes \mathbf{x}(0) = \lambda_D^k \otimes \mathbf{x}. \tag{13}$$

So the mean waiting time at crossing stops of unit intensity of passengers flow has in our system the value $w(\mathbf{A})$, which

$$w(\mathbf{A}) = \frac{\sum_{i \in \underline{n}} \sum_{j \in \underline{n}: a_{ij} > \varepsilon : \beta_j = \alpha_i \in S_c} \big(\lambda_D + x_i - x_j - a_{ij}\big)}{\sum_{i \in \underline{n}} \sum_{j \in \underline{n}: a_{ij} > \varepsilon : \beta_j = \alpha_i \in S_c} 1}. \tag{14}$$

It is possible to show that matrix $\mathbf{A}$ defined by (10) is irreducible and also from Theorem 2 only one λ_D exists that can be computed as maximal average cycle weight $\lambda(\mathbf{A})$. If eigenvector associated with λ_D is fractional number, then it is inappropiate for any real timetable. In this case we can solve

$$\mathbf{A} \otimes \mathbf{y} = \mu \otimes \mathbf{y}, \tag{15}$$

where $\mathbf{y}$ is approximate integer eigenvector for given approximate eigenvalue μ. Corresponding approximate mean waiting time of crossing stops can be calculated analogically to the form (14).

Note when we set $\mu > \lambda_D$ then we can use difference $\mu - \lambda_D$ to remove differences betweenwaiting time for transfering passengers while solving this optimization problem:

$$\sum_{i \in \underline{n}} \sum_{j \in \underline{n}: a_{ij} > \varepsilon : \beta_j = \alpha_i \in S_c} (\mu + x_i - x_j - a_{ij})^2 \quad \rightarrow \quad \min \tag{16}$$

$$\mathbf{A} \otimes \mathbf{x} \quad \leq \quad \mu \otimes \mathbf{x} \tag{17}$$

$$\mathbf{x} \quad \text{is} \quad \text{integer.} \tag{18}$$

Remark 3. For solving these problems same conditions as described in **Remark 2** apply with additional integer requirement for solution. The question is if simplier method for solving this than method of mixed quadratic programming exists.

The computational experiments are motivated by a problem of real removing one of three bus lines with one crossing stop. We begin with construction of matrix $\mathbf{A}$ for simple instance with 2 lines where every line is served by one bus. After solving corresponding eigenproblems and solutions we can create timetable for every stop of lines with these eigenvectors. There are different ways of constructing matrix $\mathbf{A}$. If we suppose first stops of buses is predetermined then length of mean period between departure times of timetable can be different. For matrix $\mathbf{A}$ defined by (10) we obtain minimal values $\lambda(\mathbf{A})$ and $w(\mathbf{A})$ but none of eigenvectors is integer. This is interesting for timetables in different peaks and troughs of day with distinct periods. This fact is not surprising for more buses on lines. The only difference against previous problem (5-6) is that solution $\mathbf{x}$ of relevant problem must be integer. Experiment done with 3 lines with 2 and 3 crossing stops by this heuristic method shows that difference $\mu - \lambda_D$) can be again used to minimize object function (16). Feasible integer solutions were found only in 83 instances of 100.

5 Conclusion

This paper introduced two models for solving the practical problems from road transport based on spectral theory of max-plus algebra. The first computational experiments of Prostějov instances [8, 10] for both max-plus models show that this approach can be applied and generates open max-plus optimization problems.

Acknowledgements

The research of authors is supported by Slovak Scientific Grant Agency under grant VEGA 1/0135/08 and by grant project SP2012/113 under Faculty of Mechanical Engineering at Technical University of Ostrava.

References

[1] Baccelli, F., Cohen, G., Olsder, G.J., Quadrat J.P.: *Synchronization and Linearity*, An Algebra for Discrete Event Systems, John Wiley & Sons, Chichester, New York, 1992.

[2] Cechlárová, K.: Eigenvectors of interval matrices over max-plus algebra, *Discrete Applied Mathematics - Special issue: Max-algebra archive*, Volume 150, Issue 1, 2005, 2 – 15.

[3] Nait-Sidi-Moh, A., Manier, M. A.; El Moudni A.: Spectral analysis for performance evaluation in a bus network, *European Journal of Operational Research*, Volume 193, Issue 1, 2009, 289 – 302, doi: 10.1016/j.ejor.2007.08.047.

[4] Heidergott, B., Oldsder, G. J., Woude, J.: *Max Plus at Work*, Modeling and Analysis of Synchronized Systems: A Course on Max-Plus Algebra and Its Applications, Princeton University Press, ISBN 0-691-11763-2, 2006.

[5] Retchkiman, K, Z.: Modeling and Analysis of the Metro-bus Public Transport System in maxixo City using Timed Event Petri nets and Max-Plus Algebra, In: *18th World IMACS/MODISIM Congres*, Carrns, Austria, 13–17 July, 2009, 1685–1991, http://mssanz.org.au/modsim09.

[6] "ScicosLab Download", ScicosLab, Retrieved 2010-04-09, http://www.scicoslab.org/

[7] Turek, M.: Využití scilabu při koordinaci světelních křižovatek, In: *Otvorený softvér vo vzdelávaní, výskume a v IT riešeniach*, Zborník príspevkov medzinárodnej konferencie OSSConf2011, 6.-9.júla 2011, Žilina, Slovensko, ISBN 978-80-970457-1-5, 2011, 111–119, in Czech.

[8] Turek, M.: Models for Coordination of Strongly Dependent Light Crossroads, PhD disertation (in Czech), Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 2012, Ostrava, Czech Republic, pp. 106.

[9] Turek, R.: Možnosti synchronizce linek MHD ve scilabu, In: *Otvorený softvér vo vzdelávaní, výskume a v IT riešeniach*, Zborník príspevkov medzinárodnej konferencie OSSConf2011, 6.-9.júla 2011, Žilina, Slovensko, ISBN 978-80-970457-1-5, 2011, 121–129, in Czech.

[10] Turek, R.: Synchronization models for the Lines of Public Transport , PhD disertation (in Czech), Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 2012, Ostrava, Czech Republic, pp. 102.

[11] De Vries, R., De Schutter, B., De Moor, B.: On max-algebraic models for transportation networks, In: *Proceedings of the 4th International Workshop on Discrete Event Systems* WODES, Cagliari, Italy, 1998, 457–462.

Effects of heavy tails on optimal investment and consumption

Jakub Petrásek [1]

Abstract. We study the effect of heavy tails on optimal proportion and consumption problem, i.e. we compare the optimal and Merton proportion and consumption and compute the wealth loss. We state and show that the effect of heavy tails is quite slight in usual conditions. The effect stays nonsignificant even if we contaminate the Lévy measure of the risky asset dynamics by severe drops of price. However, we observe that heavy tails need to be taken into account if an investor is exposed to a very huge loss or even bancruptcy. This could be the case of a very aggressive investor. Finally, we study the lower bound of the optimal investment proportion. We show that even for infinite kurtosis the optimal investment proportion is still positive. We also studied the rate of convergence to zero of the optimal investment proportion as the volatility/risk averse coefficient approaches infinity or expected return approaches zero.

Keywords: Optimal portfolio and consumption, Jump processes

JEL classification: G11
AMS classification: 91G80

1 Introduction

Optimal consumption and allocation of wealth between a risky asset and a bank account is a crucial question for any investor. This problem dates back to Merton [6], who models the risky asset by geometric Brownian motion. However, empirical tests reveals that the distribution of relative asset returns may be far from gaussian. They are characteristic by heavy tails and asymmetry. If we allow jumps in the model, we obtain heavy tailed with possibly asymmetric distribution. Note also that it is natural not to assume that the trajectory of asset price is continuous because the information which influence the price reveals discontinously.

The theoretical formulas for the optimal investment and consumption strategy for a jump-diffusion process and under CRRA utility function were derived in [5]. Oksendal in [8] in example 3.1 shows that adding jumps decreases the optimal proportion. The effect on optimal consumption differs for aggressive (with risk averse coefficient lower than 1) and conservative investors. The aggressive would consume smaller proportion relative to Merton while the conservative bigger. Cvitanić [4] performs a numerical study and suggests a Variance Gamma process as a reasonable model for risky asset price dynamics. A different model for the risky asset price dynamics, Normal Inverse gaussian process, is used in [9], where the optimal consumption proportion is also empirically studied. Benth et al. in [2] calibrates Normal Inverse Gaussian to the financial data and then applies the optimal control for different values of risk free rate.

The numerical study in [9] indicates that the effect of jumps is very slight for usual values of skewness and kurtosis. Furthermore, the effect is still quite insignificant for extremely high kurtosis and becomes significant for a very negatively skewed asset returns. In the following paper we show that the effect of heavy tails increases rapidly as the investor approaches 'the bankruptcy region'. We also prove that even for infinite kurtosis the optimal proportion is still positive. Furthermore, we study limits of the optimal relative to Merton proportion as moments and risk averse coefficient approach boundary values.

The paper is organized as follows. In the section 2 the economic model and its solution are presented.

[1]Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University in Prague, e-mail: petrasek@karlin.mff.cuni.cz

Section 3 deals with numerical application. The results of paper Petrásek [9] are briefly described and extended by a model with additional source of risk. Finally, section 4 is devoted to the study of lower bound of the investment proportion and its consequences are discussed.

2 Theoretical Model

Consider an investor placing his money into two assets, risk-free, paying interest rate r, and an risky asset that follows a geometric Lévy process

$$\mathrm{d}\,S_t = S(t-)\left(\alpha\,\mathrm{d}\,t + \sigma\,\mathrm{d}\,W_t + \int_{-1}^{\infty} z\widetilde{N}(\mathrm{d}\,t, \mathrm{d}\,z)\right), \quad t > 0, \tag{1}$$

where W is Brownian motion and $\widetilde{N}$ is process of compensated Poisson random measure, both on a filtered probability space $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P\right)$.

At any time $t \geq 0$ an investor can choose the number Δ_t of risky assets S_t in his portfolio and he can also consume money from his account at rate $C_t \geq 0$. Let further $\theta_t = \frac{\Delta_t S_{t-}}{X_{t-}}$ denotes the proportion of his capital invested in risky asset at time t and $c_t = \frac{C_t}{X_{t-}}$ the consumption proportion. Then the value of his portfolio X_t is given by the stochastic differential equation

$$\begin{aligned} \mathrm{d}\,X_t &= \Delta_t\,\mathrm{d}\,S_t + rX_t\,\mathrm{d}\,t - C_t\,\mathrm{d}\,t & (2)\\ &= X_{t-}\theta_t\left(\alpha\,\mathrm{d}\,t + \sigma\,\mathrm{d}\,W_t + \int_{-1}^{\infty} z\widetilde{N}(\mathrm{d}\,t, \mathrm{d}\,z)\right) + rX_t\,\mathrm{d}\,t - c_tX_t\,\mathrm{d}\,t, & (3)\end{aligned}$$

with $X(0) = x$, $\theta_t, c_t \in \mathcal{F}_{t-}$ (predictable).

Definition 1. An $\mathcal{F}_{t-}$ adapted policy (Δ_t, C_t) is admissible if $C_t \geq 0$ and X_t, $t \geq 0$ is the wealth and consumption portfolio (3). We denote $\mathcal{A}(x)$ a set of admissible policies with $X_0 = x$.

The objective of an investor is to maximize his discounted consumption C_t, i.e. we seek a function v (called a *value function*) such that

$$v(x) = \sup_{(\theta_t, c_t)\in\mathcal{A}(x)} \int_0^{\infty} e^{-\beta t}\,\mathrm{E}\,U(C_t)\,\mathrm{d}\,t, \tag{4}$$

where β is a discount factor and U denotes a power utility function of the form

$$\begin{aligned} U(x) &= \frac{x^{1-p}}{1-p}, \quad p > 0, p \neq 1,\\ &= \log(x), \quad p = 1.\end{aligned}$$

Theorem 1 (Optimal Consumption and Portfolio)**.** *Assume the portfolio (3) and the objective (4) and assume that the risk aversion coefficient $p \neq 1$. Let us denote*

$$\Lambda(\theta) = \alpha_t\theta(1-p) - \frac{1}{2}\sigma_t^2\theta^2 p(1-p) + \int_{-1}^{\infty}\left((1+\theta z)^{1-p} - 1 - \theta z(1-p)\right)\nu_t(dz)$$

and let θ^* be such that

$$\Lambda'(\theta^*) = 0. \tag{5}$$

Assume also that

$$\beta - r(1-p) - \Lambda(\theta^*) > 0.$$

Then θ^* is the optimal investment proportion, $c^* = (K(1-p))^{-1/p}$ is the optimal consumption and $v(z) = Kz^{1-p}$ is the value function, where $K = \frac{1}{1-p}\left(\frac{\beta - r(1-p) - \Lambda(\theta^*)}{p}\right)^{-p}$.

Proof. The proof is based on using the dynamic programming principle and solving the acquired Hamilton-Jacobi-Bellman equation. For a general theorem (Optimal Control for Jump Diffusions) see [8], theorem 3.1 and subsequent example 3.2 or [2]. □

The theorem solves the investors problem, however, leads to an integro-differential equation, which must be solved numerically in general.

3 Empirical results

The empirical study performed in [9] shows quite slight effect of heavy tails. The numerical study proceeded as follows. The process of logarithmic prices L_t of the risky asset was modeled by a Normal Inverse Gaussian process, see [1], i.e. it can be described by a pure jump Lévy process of the form

$$\mathrm{d}\, L_t = b_L \,\mathrm{d}\, t + \int_{\mathbb{R}} z \widetilde{N}_L(\mathrm{d}\, t, \mathrm{d}\, z), \quad 0 \leq t < \infty,$$

where $\widetilde{N}_L$ is a compensated Poisson random measure with Lévy measure ν_L. Applying Itô formula we get the dynamics of the prices (1)

$$\mathrm{d}\, S_t = \mathrm{d}\exp(L_t) = S_{t-} \left\{ \left(b_L + \int_{\mathbb{R}} (e^z - 1 - z)\, \nu_L(z) \right) \mathrm{d}\, t + \int_{\mathbb{R}} (e^z - 1)\, \widetilde{N}_L(\mathrm{d}\, t, \mathrm{d}\, z) \right\}.$$

Finally, the equation (5) was solved.

The numerical study aimed at the effect of skewness and kurtosis on the optimal proportion relative to the Merton proportion θ_0. Let us fix the risk aversion coefficient $p = 6$, volatility $\sigma = 16$ % (p.a.), Sharpe ratio of the risky asset as 0.40, risk free rate 2 % (p.a.) and discount parameter β 10 %. Then the optimal proportion relative to the Merton proportion was more than 99.5 % for symmetric returns with a very high excess kurtosis (30). The effect of skewness was much stronger with decrease of the optimal proportion of almost 2 % for skewness level equal to -1. The results are summed up in table 1.

We extend the study by computing a *wealth loss* w. Wealth loss is the initial added wealth the investor ignoring jumps in the model requires such that he has the same utility from discounted consumption as if he allocates and consumes his money optimally, i.e. we solve the following equation[1]

$$v\left(x(1+w); \theta_0, c_0\right) = v(x; \theta^*, c^*),$$

where $v(x, \theta, c)$ is the value function for a given strategy (θ, c) with initial wealth x. We can see that an investor ignoring jumps requires only additional \$1.4 per \$10000 for negatively skewed relative returns and excess kurtosis 30.

σ	κ_3	κ_4	θ^*/θ_0	c^*/c_0	w
0.160	-1.000	30.000	98.096	99.700	0.000143
0.160	-0.000	30.000	99.510	99.939	0.000008
0.160	1.000	30.000	101.008	100.188	0.000040
0.160	-1.500	60.000	96.943	99.522	0.000381
0.160	-0.000	60.000	99.028	99.877	0.000034
0.160	1.500	60.001	101.312	100.256	0.000069

Table 1: Optimal consumption and investment proportion for different values of skewness and kurtosis. κ_3 denotes skewness and κ_4 kurtosis of asset returns.

An investor must also face to an unobservable effects. These are the events that appears with very low frequency, and so are left out from the model[2]. We conclude that our model lacks severe drops of asset price. If we assume, that such events can be described by an independent Lévy process, we can use the fact that if we have two independent Lévy processes they never jump at the same time (corollary of Theorem 4.1 and Theorem 5.1 in [3]). In other words if we have two independent processes $\{L_t\}$ and $\{E_t\}$ with Lévy measures ν_L and ν_Y then the Lévy measure of their sum is defined as

$$\nu_X(B) = \nu_L(B) + \nu_E(B), \quad \forall B \in \mathbb{R}\,.$$

[1] Another possibility how to compare optimal and suboptimal strategy is to compute the certainty equivalence for both strategies.

[2] Generally, models assume stationary time series. Thus they are calibrated using quite a short (stationary) time period.

We assume that the process of unobservable events $\{E_t\}$ can be modeled by a Poisson compound process with a very low intensity and gaussian jumps[3], i.e.

$$E_t = \sum_{k=0}^{N_t} Y_k - \lambda \mu_E t, \quad Y_k \sim N(\mu_E, \sigma_E^2),\ N_t \sim Po(\lambda t).$$

Parameters are set so that the intensity of such events is every four years and the day relative loss is approximately 10 %[4]. The Lévy measure of this process is given by the formula

$$\nu_E(x) = \lambda \cdot \frac{1}{\sqrt{2\pi}\sigma_E} \cdot \exp\left(-\frac{(x-\mu_E)^2}{2\sigma_E^2}\right), \quad x \in \mathbb{R}\,. \tag{6}$$

We performed a similar numerical study as previously, but with additional source of risk, given by the process of unobservable events $\{E_t\}$. We found that the effect of additional jumps is almost negligible, see table 2. The contamination reduces the investment proportion into the risky asset by about 0.5 % and the wealth loss is still only about \$2.4 per \$10000 invested for negatively skewed relative returns. We can conclude that the effects of higher moments are not so serious in usual conditions which is consistent with Cvitanić [4]. Cvitanić [4] also observed, that increasing volatility with Merton proportion kept fixed increases the difference between optimal and Merton investment proportion. Note that, it is equivalent with increasing the Sharpe ratio, whose strong effect was observed in [9]. In the following section we will study the effect of risk aversion coefficient.

σ	κ_3	κ_4	θ^*/θ_0	c^*/c_0	w
0.160	-1.227	33.219	97.582	99.612	0.000236
0.160	-0.227	33.219	99.007	99.856	0.000037
0.160	0.773	33.219	100.474	100.102	0.000010
0.160	-1.727	63.219	96.482	99.446	0.000511
0.160	-0.227	63.219	98.546	99.797	0.000080
0.160	1.273	63.220	99.814	99.929	0.000003

Table 2: Optimal consumption and investment proportion for different values of skewness and kurtosis. κ_3 denotes skewness and κ_4 kurtosis of asset returns.

4 Poisson type models

We will use compound Poisson models possibly with diffusion part. Note that as long as jumps are allowed in the model, the investment proportion cannot exceed the level $1/J$, where J is the biggest possible drop of price, i.e. if the model allows 50% loss of the risky asset price our investment proportion cannot exceed 2. On the other hand, jumps (for moderate values of volatility, skewness and kurtosis) reduce the optimal investment proportion only by a several percent.

Let us assume the following model for asset price dynamics

$$\mathrm{d}\, S_t = S_{t-}\left(\alpha\, \mathrm{d}\, t + \sigma_C\, \mathrm{d}\, W_t + J\, \mathrm{d}\, \widetilde{N}_t\right), \tag{7}$$

where J is the given jump size and $\widetilde{N}_t$ is a compensated Poisson process with intensity λ. Thus the volatility of relative returns is $\sigma = \sigma_C + \lambda \cdot J^2$. We set $\lambda = 0.0001$ (per day), jump -0.3, then skewness equals -2.605 and excess kurtosis equals 77.248.

In picture 1 we can see that the effect of added jump is almost negligible for a conservative investor with risk aversion coefficient higher than 3 and becomes extremely significant as the risk aversion decreases below the value 1. Jumps significantly influence the optimal proportion as well as the optimal consumption as the investor approaches the border of solvency region. The optimal relative to Merton consumption is higher for aggressive investor with $p < 1$.

[3] Poisson compound process with gaussian jumps plus diffusion part is called Merton model, see [7].

[4] $\lambda = 0.001$ (per day) $\mu_E - 0.1$ (per day), $\sigma_E = 0.05$ (per day).

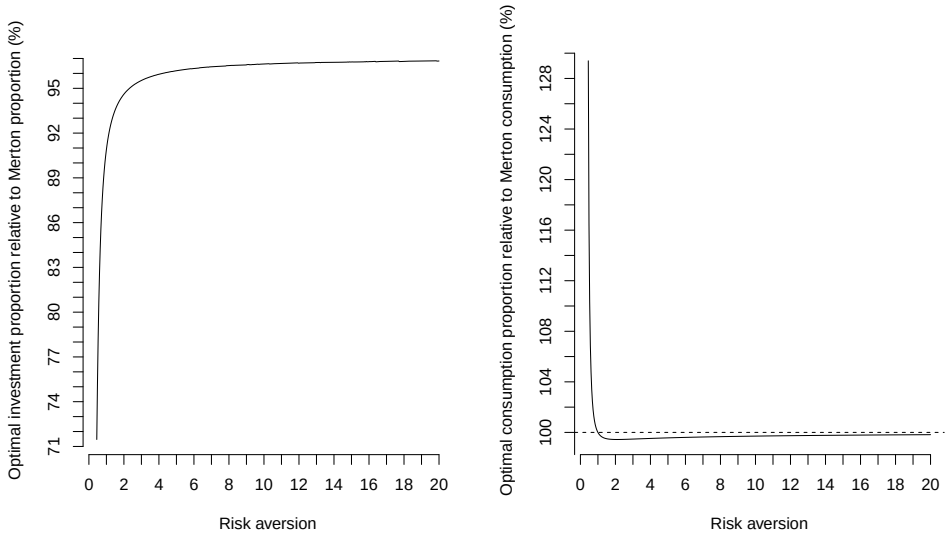


Figure 1: Optimal proportion for model (7) relative to Merton proportion and optimal consumption proportion for model (7)relative to Merton proportion consumption as a function of risk aversion coefficient.

We can ask whether there exist a distribution of risky asset returns that leads to zero optimal proportion. It is enough to think about the worst scenario for the investor, where the risky asset follows a compound Poisson process with a Dirac measure at 100% loss. The Lévy measure of the described model is given by the formula

$$\nu(z) = \lambda \cdot \delta_{-1}(z),$$

where δ_x denotes a Dirac measure at point x and λ, which determines variance σ^2, is set such that the volatility is still 16 % (p.a.). If we plug the proposed Lévy measure into the equation (5) we see that the optimal proportion solves the equation

$$\alpha - \sigma^2 \cdot \left(\frac{1}{(1-\theta)^p} - 1 \right) \quad = \quad 0.$$

Proposition 2. *The lower bound for the investment proportion is*

$$\theta_{LB} = 1 - \left(1 + \frac{\alpha - r}{\sigma^2} \right)^{-1/p}.$$

Note that we found the optimal proportion lower bound for any risky asset model of the type (1) (geometric Lévy process), thus also for models with infinite kurtosis.

Proposition 3. *Assume a model, where the only risk source is* 100*% loss. Then*

$$\begin{aligned} \frac{\theta_{LB}}{\theta_0} &\to 1 \quad as \quad \frac{\alpha - r}{\sigma^2} \to 0, \\ \frac{\theta_{LB}}{\theta_0} &\to \frac{\sigma^2}{\alpha - r} \cdot \log\left(\frac{\alpha - r}{\sigma^2} + 1 \right) \quad as \quad p \to \infty, \end{aligned}$$

for all other parameters being fixed.

Proof. Taylor expansion in $\frac{\alpha - r}{\sigma^2}$ of the first order is

$$\theta_{LB} = \frac{\alpha - r}{p\sigma^2} + o\left(\frac{\alpha - r}{\sigma^2} \right)$$

and in $1/p$, we obtain

$$\theta_{LB} = \frac{1}{p} \log\left(\frac{\alpha - r}{\sigma^2} + 1 \right) + o\left(\frac{1}{p} \right).$$

□

We conclude that if a risky asset follows a geometric Lévy process. Then

$$\frac{\theta^*}{\theta_0} \to \gamma \quad \text{as} \quad \frac{\alpha - r}{\sigma^2} \to 0,$$

where $\gamma \geq 1$ and the rate of convergence (to zero) of θ^* and θ_0 as $p \to \infty$ differs.

5 Conclusion

In this paper we studied the effect of heavy tails on optimal investment proportion and consumption. We continued in the study by Petrásek [9]. We extended the results by adding a severe drops of price to the original risky asset price dynamics. We conclude that the influence of heavy tails is still very slight. However, it may be significant for an aggressive investor or if the risky asset has a very high Sharpe ratio. We observed that as an aggressive investor reaches the borders of solvency region the optimal relative to Merton investment proportion decreases very rapidly while the optimal relative to Merton consumption proportion increases. Furthermore, we included the *wealth loss* in the numerical study and inferred that a suboptimal investor (using Merton proportion and consumption under heavy tailed and possibly negatively skewed risky asset returns) loses only a few money units per 10 000 units invested.

Finally, we derived the lower bound for the investment proportion. We proved that even for infinite kurtosis the optimal proportion is still positive. Furthermore, we proved that as the excess expected return of the risky asset approaches zero or its volatility approaches infinity the optimal proportion does converge to zero at the same rate as Merton investment proportion. Thus Merton investment proportion can serve as a good approximation in such cases. But as the risk aversion coefficient goes to infinity the rate of convergence generally differs. Hence if an investor holds a very small proportion in the risky asset due to his high risk aversion, he cannot neglect jumps in general.

Acknowledgements

The work was supported by the grant SVV 261315/2012 and by the grant 17510 of the Grant Agency of Charles University in Prague.

References

[1] Barndorff-Nielsen, O. E.: Processes of Normal Inverse Gaussian Type. *Finance and Stochastic* **2** (1998), 41–68.

[2] Benth, F. E. and Karlsen, K. H. and Reikvam, K.: A note on portfolio management under non-Gaussian logreturns. *Int. J. Theor. Appl. Finance* **5** (2001), 711–731.

[3] Cont, R. and Tankov, P.: *Financial modelling with jump processes.* Chapman & Hall/CRC Financial Mathematics Series, 2004.

[4] Cvitanić, J. and Polimenis, V. and Zapatero, F.: Optimal portfolio with higher moments. *Int. J. Theor. Appl. Finance* **4** (2008), 1–28.

[5] Framstad, N. C. and Øksendal, B. and Sulem, A.: Optimal consumption and portfolio in a jump diffusion market. In: *A. Shyriaev et al (eds): Workshop on Mathematical Finance*, 1998, 9–20.

[6] Merton, R. C.: Optimum consumption and portfolio rules in a continuous-time case. *J. Econ. Theory* **3** (1971), 373–413.

[7] Merton, R. C.: Option pricing when underlying stock returns are discontinuous. *J. Financ. Econ.* **3** (1976), 125–144.

[8] Øksendal, B. and Sulem, A.: *Applied stochastic control of jump diffusions.* Universitext. Berlin: Springer, 2007.

[9] Petrásek, J.: Optimal portfolio and consumption under pure jump process. In: *Proceedings of Conference on Mathematical Methods in Economics 2010*, České Budějovice, 2010, 517–522.

Behavioral equilibrium exchange rate in Greece and Ireland

Klára Plecitá[1], Luboš Střelec[2]

Abstract. Open macroeconomics sees real exchange rate as an important tool in maintaining macroeconomic balance. When euro area countries adopted the common currency euro, they have lost the opportunity to adjust their real exchange rates via changes in nominal exchange rates, thus these countries can correct for real exchange rate overvaluation only through internal devaluation or an exit from the euro area. This paper focuses on equilibrium exchange rates for Greece and Ireland since their entrance into the euro area until the end of the year 2011. To examine the real exchange rate misalignment for Greece and Ireland towards the rest of the euro area, the concepts of Behavioral Equilibrium Exchange Rate (BEER) and Permanent Equilibrium Exchange Rate (PEER) are applied.
We find that the real exchange rate of both Greece and Ireland was undervalued towards the euro around the time, when these countries entered the euro area. Further we uncover that for the rest of the period analyzed the Irish and Greek real exchange rate was broadly in line with underlying fundamentals.

Keywords: equilibrium exchange rate, BEER, PEER, Ireland, Greece.

JEL Classification: C32, C51, C58, F31
AMS Classification: 91G70, 91G80

1 Introduction

According to open macroeconomics, real exchange rate is a very important tool in attaining internal and external macroeconomic balance. When euro area countries adopted the common currency euro, they have lost the opportunity to adjust their real exchange rates via changes in nominal exchange rates. Therefore it is of high importance that countries enter the euro area with properly set conversion rates of their national currencies to the euro. If a country joins the euro area with an overvalued currency it may undermine competitiveness of its businesses and lead to lower exports with a negative impact on economic growth. Similarly, joining the euro area with an undervalued currency may boost competitiveness and lead to an overheating of the economy. Therefore, persistent real exchange rate misalignments may lead to macroeconomic imbalances and resources misallocation, and elimination of these imbalances, especially in the context of the euro area, which consists of a set of sovereign countries with not fully synchronized economic and political course of events, may be done via internal devaluation or revaluation process or an exit from the euro area. In recent years not only the world macroeconomic imbalances but also the imbalances within the frame of the euro area are widely discussed. Plecitá and Střelec [7] identified that the largest macroeconomic imbalances in the euro area in the year 2010 displayed Greece and Ireland. Therefore the aim of this paper is to estimate the equilibrium real effective exchange rate for Greece and Ireland towards the rest of the euro area (which is for our purpose defined as the original EA-12) and consequently to compute the degree of their exchange rate misalignment.

To assess the degree of exchange rate misalignment, which is defined as a gap between actual (observed) real effective exchange rate and equilibrium exchange rate, it is necessary to estimate the equilibrium exchange rate, for which several concepts are available. The traditional theory of the equilibrium exchange rate is the Purchasing Power Parity (*PPP*) which explains the movements in equilibrium exchange rates in the long run on the basis of relative prices. Beyond the *PPP*, two other major equilibrium exchange rate theories and their various modifications were developed: the Fundamental Equilibrium Exchange Rate (*FEER*) and the Behavioral Equilibrium Exchange Rate (*BEER*). The equilibrium exchange rate in *FEER* is defined as the exchange rate that simultaneously ensures internal and external macroeconomic balance [11]. The internal equilibrium is reached when economy is operating at full employment in a low inflation environment and the external equilibrium is characterized by a sustainable current account. However defining the sustainable position of the current account is highly controversial and subjective in nature [10]. In contrast, in the *BEER* approach the equilibrium exchange rate is consistent with prevailing level of economic fundamentals and rests on the principles of uncovered interest rate parity [1]. A comprehensive discussion of the different approaches for estimating equilibrium exchange rate can

[1] MENDELU, Department of Finance, Zemědělská 1, 613 00 Brno, Czech Republic, klara.plecita@mendelu.cz.
[2] MENDELU, Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, lubos.strelec@mendelu.cz.

be found for example in MacDonald [5]. As the *BEER* approach is said to be suitable for relatively short time series [8], as is the case of our euro-area time series, we decided to estimate the equilibrium exchange rate for Greece and Ireland during their membership in the euro area using the *BEER* methodology.

2 Data and methodology

The behavioral equilibrium real effective exchange rate for Greece and Ireland is estimated towards the rest of the original 12 euro area countries (EA-12) on the quarterly basis from the entrance of these countries into the euro area (Ireland 1999 and Greece 2001) until the end of the year 2011. The empirical estimations of the *BEER* model in this paper draw heavily on Clark and MacDonald [1]. The *BEER* approach produces estimates of equilibrium real exchange rate which incorporate both short run and long run economic fundamentals. Based on the literature, the *BEER* is determined by following set of economic variables:

$$BEER = f\left((r - r^*), TNT, TOT, IIP, DEBT\right) \tag{1}$$

where *BEER* is the behavioral equilibrium real exchange rate estimated on the basis of the real effective exchange rate (*REER*) of the domestic economy to the rest of the EA-12 countries. *REER* is computed from trade-weighted nominal effective exchange rate deflated by the ratio of domestic and foreign price indexes (HICP). Data source: European Commission.

(r - r^*) is the real interest rate differential that represents the uncovered interest rate parity, r is the domestic real interest rate and r^* is the foreign real interest rate, i.e. the weighted real interest rate for the rest of the EA-12 countries. An increase in the real interest rate differential is expected to lead to an appreciation of the real exchange rate. The real interest rates were calculated using quarterly nominal interest rates for 10-year government bonds, which were deflated using HICP. Data source: Eurostat.

TNT is the relative price of non-traded goods in domestic economy relative to the price of traded goods in domestic economy relative to the equivalent of foreign effective (trade weighted) foreign ration of non-traded to traded goods. The *TNT* variable is used as a productivity proxy stemming from the Balassa-Samuelson effect and measures the impact of productivity on the real exchange rate. An increase in the *TNT* should lead to appreciation of the real exchange rate. The *TNT* was computed as a ratio of consumer price index (HICP) and producer prices in manufacturing (PPI). Data source: OECD Stat.

TOT represents terms of trade of domestic economy relative to trade-weighted terms of trade of foreign economies. An increase in *TOT* should improve trade balance and therefore lead to appreciation of real exchange rate. The terms of trade were computed as the ratio of the export to import indexes. Data source: Eurostat.

IIP stands for net international investment position, which is determined by the attitude of the public and private sector towards savings and investment. The impact of *IIP* on real exchange rate is ambiguous, i.e. increase in *IIP* can lead either to appreciation or depreciation of real exchange rate depending on other circumstances (for more details see Maeso-Fernandez et al. [6]). *IIP* is computed as a ratio of the stock of net international investment position to nominal gross domestic product. As data on *IIP* are prior the year 2004 available only on yearly bases, the data for the time period 1Q1999 – 4Q2003 were interpolated from the yearly data using cubic spline interpolation method. Data source: IFS IMF.

DEBT is the ratio of domestic general government debt to general government debt of the rest of the EA-12 countries. *DEBT* is used as a proxy for time variant risk. An increase in *DEBT* leads to depreciation of the real exchange rate. Data source: Eurostat.

Econometric estimation of *BEER* is based on linear semi-logarithmic form expressed as

$$\log REER_t = \alpha_0 + \alpha_1 IIP_t + \alpha_1 (r - r^*)_t + \alpha_2 \log TOT_t + \alpha_3 \log TNT_t + \alpha_4 \log DEBT_t + \varepsilon_t \tag{2}$$

where log denotes a logarithm and ε_t is the error term.

Before applying the Johansen (1995) cointegration method, stationarity of the variables analyzed is tested via the Augmented Dickey-Fuller (ADF) unit root test, because at least some (but not necessarily all) of the variables entering the Johansen cointegration method must be nonstationary [1]. And cointegration vectors on the basis of the Trace test statistic are identified. In the next step equation (2) is estimated and statistically insignificant coefficients are removed by applying sequential elimination methods. Consequently, these restricted models for selected countries are used for estimation of the behavioral equilibrium exchange rate. Moreover two types of equilibrium exchange rates are computed: the medium-run *BEER* and the long-run *BEER* called *PEER* (permanent equilibrium real exchange rate). The medium-run *BEER* values are estimated, when current (actual) values of explanatory variables are used in the cointegration relationship. It is, however, possible, that the current values

of economic fundamentals may depart from their sustainable levels during period analyzed. Therefore the *PEER* values are estimated using equilibrium values of explanatory variables, which are obtained from the Hodrick-Prescott filter [3].

In the last step current and total misalignments of the real effective exchange rates are computed. Current misalignment is given by the current values of explanatory variables and is obtained as the difference between actual *REER* and estimated equilibrium *BEER* values. Total misalignment is defined as a gap between actual *REER* and estimated lung-run *PEER* values.

3 Equilibrium exchange rates

Unit root tests, which results can be seen in Table 1, applied to the data suggest that all series are nonstationary at 1 % significance level, and most of them are integrated of the order one. Therefore, we applied the Johansen cointegration method. On the basis of the Trace statistics (Table 2), there would appear to be up to four cointegration vectors in the case of Greece and up to 3 cointegration vectors for Ireland.

	Greece		**Ireland**	
	statistic	**p-value**	**statistic**	**p-value**
log *REER*	-2.574	0.292	-1.668	0.766
IIP	-1.256	0.898	-2.576	0.292
(*r*-*r**)	2.351	1.000	-2.059	0.568
log *TOT*	-2.232	0.471	-3.768	0.018
log *TNT*	-1.951	0.627	-2.263	0.454
log *DEBT*	-1.643	0.776	-1.817	0.697

Table 1 ADF unit root tests for variables in *BEER* models

	Greece		**Ireland**	
$H_0 : r$	**Trace**	**p-value**	**Trace**	**p-value**
0	191.9	0.000	151.5	0.000
1	136.2	0.000	90.5	0.000
2	86.1	0.000	54.5	0.009
3	43.3	0.001	25.4	0.150
4	14.9	0.060	9.7	0.308
5	0.3	0.559	1.7	0.191

Table 2 Trace test statistics of cointegration vectors

As is known, the existence of multiple cointegration vectors complicates the interpretation of equilibrium; therefore we restricted the cointegration space only to one cointegration vector for log *REER* for each country (see Table 3).

	Normalized cointegration coefficient						
	log *REER*	*IIP*	(*r*-*r**)	log *TOT*	log *TNT*	log *DEBT*	const.
Greece	1.000	0.045	-0.005	0.047	0.230	0.011	-0.005
	(0.000)	(0.004)	(0.001)	(0.057)	(0.093)	(0.034)	(0.955)
Ireland	1.000	0.018	0.005	0.167	-0.152	0.029	-0.131
	(0.000)	(0.002)	(0.001)	(0.085)	(0.019)	(0.008)	(0.087)

Table 3 Cointegration analysis and EC model (standard errors in parenthesis)

Some of the coefficients estimated were found to be statistically insignificant; therefore in the next step we used sequential elimination methods to estimate the country-specific restricted models (see Table 4).

	Greece		Ireland	
	coefficient	**p-value**	**coefficient**	**p-value**
const.	4.576	0.000	4.613	0.000
IIP	-0.035	0.000	-0.009	0.000
(*r*-*r**)	0.001	0.001	0.001	0.013
log *TOT*				
log *TNT*	-0.275	0.001	0.091	0.000
log *DEBT*			-0.030	0.000

Table 4 Estimates of log *REER* (restricted model)

All of the variables entering the restricted model (*BEER*) for Greece (Table 4) are statistically significant at 5 % significance level and have the correct sign, apart from the *TNT* sign, which is negative. The *TNT* is an imperfect measure of relative non-traded prices and can be affected by various factors unrelated to the Balassa-Samuelson framework (for more detailed discussion see for example Engel [2] and Schnatz et al. [9]); therefore Rusek [8] argues that the sign of *TNT* can be both positive and negative and depends on the data. Hence we did not exclude this variable from our estimated model. All of the variables entering the restricted *BEER* model for Ireland (Table 4) are statistically significant at 5 % level and have the expected sign.

In order to generate the long-run *PEER* for Greece and Ireland, smoothed values of explanatory variables (from the Hodrick-Prescott filter) were substituted into the estimated *BEER* equations (Table 4). Figure 1 plots the *BEER* and *PEER* estimates for Greece and Ireland alongside with their actual *REER* values during the period 1Q2001 – 4Q2011, respectively 1Q1999 – 4Q2011.

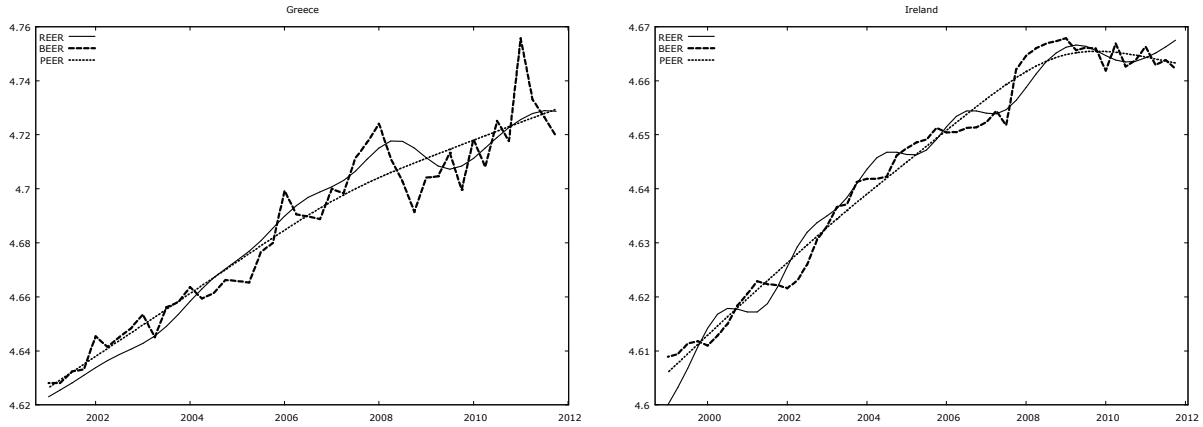


Figure 1 Plot of the (logarithm of the) actual *REER*, medium-run *BEER* and long-run *PEER* for Greece and Ireland to the rest of the EA-12

As can be seen from Figures 1 and 2, the difference between the actual national real effective exchange rate (*REER*) of Greece and Ireland and their medium-run (*BEER*) and long-run equilibrium values (*PEER*) are neither large nor persistent. Both countries entered the euro area with slightly undervalued *REER*s, as can be seen from Figure 2 - both on the basis of current and total misalignment, but this undervaluation was corrected, especially in the case of Ireland, very shortly afterwards. The Irish actual *REER* oscillated very closely around its equilibrium *BEER* and *PEER* values. In the case of Greece the periods of misalignment of actual *REER* from its long-term equilibrium *PEER* values were more prolonged, but still they were hardly ever larger than one percentage point.

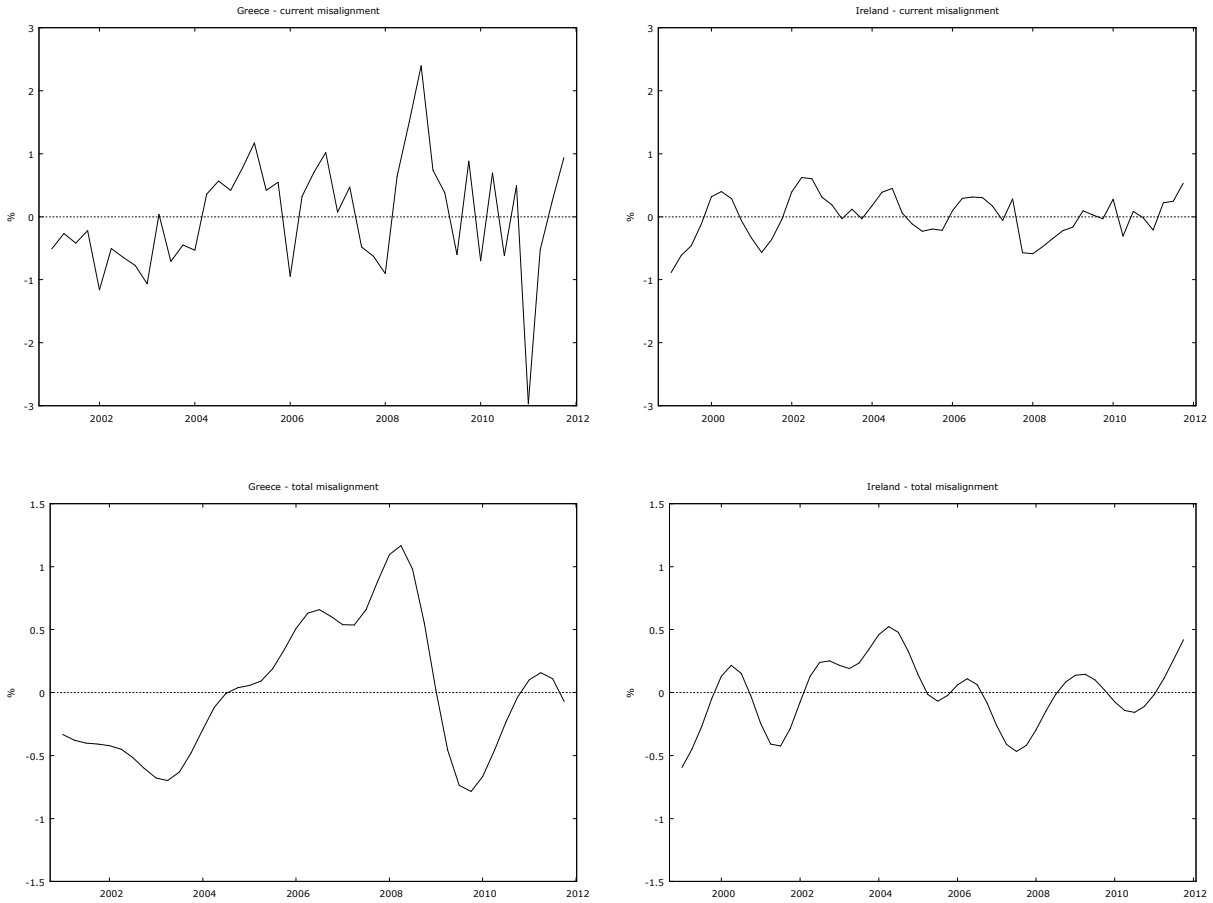


Figure 2 Current and total misalignments for Greece and Ireland to the EA-12

For Greece's membership in the euro area large and persistent current account deficits, debt accumulation and rise in labor costs have been typical. And the Irish economy was severely hit by the economic crisis. Therefore the results from our *BEER* and *PEER* analysis, which suggest, that the Greek and Irish *REERs* have been more or less consistent with their underlying fundamentals over period analyzed, could be striking. However it must be noted that we focus only on the Greek equilibrium towards the original euro area, which consisted only of 12 members states, therefore the Greek and Irish equilibrium exchange rate towards the rest of the world may look different. Moreover, as noted by Siregar and Rajan [10], the *BEER* and *PEER* equilibriums are from definition not consistent with the typical notion of macroeconomic balance, i.e. simultaneous achievement of internal and external equilibrium, but they are consistent with prevailing levels of macroeconomic fundamentals, respectively with their equilibrium levels.

Finally, there is scope for additional research in this area. As noted in the beginning of this paper there are different models for estimating equilibrium exchange rates, which are based on diverse sets of variables and economic assumptions, and this paper is based only on the *BEER* approach. To enable us to draw more robust conclusions about the equilibrium real exchange rates for Greece and Ireland during their membership in the euro area we need to apply also the other approaches which will be the focus of our further research.

4 Conclusion

The aim of this paper was to estimate equilibrium real effective exchange rate for Greece and Ireland during their membership in the euro area and to compute the degree of their exchange rate misalignment toward the rest of the original euro area (EA-12). To estimate the equilibrium real exchange rate the behavior equilibrium exchange rate (*BEER*) promoted by Clark and MacDonald [1] has been applied. Moreover to eliminate the effect of temporary shocks on the equilibrium exchange rate the permanent equilibrium exchange rate (*PEER*) has been computed. In particular, we find that the actual real effective exchange rates (*REER*) of Greece and Ireland towards the rest of the EA-12 are broadly in line with their equilibrium *BEER* and *PEER* levels. Both Greece and

Ireland entered the euro area with an undervalued *REER*, but this misalignment has been corrected afterwards, and during the period analyzed there has been found no persistent period of misalignment.

Acknowledgements

This paper is a part of outputs of research intention supported by the Ministry of Education, Youth and Sports of the Czech Republic (Project No. MSM 6215648904/01).

References

[1] Clark, P. B., and MacDonald, R.: *Exchange Rates and Economic Fundamentals – A Methodological Comparison of BEERs and FEERs*. IMF Working Paper WP/98/67, International Monetary Fund. 1998.

[2] Engel, C.: Accounting for U.S. Real Exchange Rate Changes. *Journal of Political Economy* **107** (1999), 507–538.

[3] 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.

[4] Johansen, S.: Identifying restrictions of linear equations with applications to simultaneous equations and cointegration. *Journal of Econometrics* **69** (1995), 111–132.

[5] MacDonald, R.: *Concepts to Calculate Equilibrium Exchange Rates: An Overview*. Frankfurt am Main: Deutsche Bundesbank, 2000.

[6] Maeso-Fernandez, F., Osbat, C., and Schnatz, B.: Determinants of the Euro Real Effective Exchange Rate: A Beer/Peer Approach. *Australian Economic Papers* **41** (2002), 437–461.

[7] Plecitá, K., and Střelec, L.: On equilibrium real exchange rates in euro area: Special focus on behavioral equilibrium exchange rates in Ireland and Greece. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* **LX-2** (2012), 261–269.

[8] Rusek, A.: Real Effective Exchange Rate for the Euro: Equilibrium and Misalignment. In: *The Economic Performance of the European Union: Issues, Trends and Policies* (Rusek, A., Lacina, L., and Fidrmuc J., eds.). Palgrave Macmillan, London, 2009, 42–65.

[9] Schnatz, B., Vijselaar, F., and Osbat, C.: Productivity and the Euro-Dollar Exchange Rate. *Review of World Economics* **140** (2004), 1–30.

[10] Siregar, R., and Rajan, R.: *Models of Equilibrium Real Exchange Rates Revisited: A Selective Review of the Literature*. Centre for International Economic Studies, University of Adelaide, 2006.

[11] Williamson, J.: *Estimating Equilibrium Exchange Rates*. Peterson Institute of International Economics, Washington DC, 1994.

The macro-financial linkages modelling for the Czech economy

Jiří Polanský[1], Jaromír Tonner[2], Osvald Vašíček[3]

Abstract. The contribution presents and analyze the model with financial frictions. It is tailor-made for the Czech economy, and thus contains several features for capturing Czech stylized facts (a cascade of nominal rigidities, high openness, real exchange rate appreciation in consumer prices etc.). Linkages between real and financial sectors are incorporated via the state non-contingent debt-contracts within the financial accelerator. Also, the model contains shocks which hit financial variables and propagate through the model into real sectors. The empirical analysis is presented via results of the Bayesian estimation.

Keywords: financial frictions, DSGE models, Bayesian methods

JEL classification: C53, E32, E37
AMS classification: 90C15

1 Introduction

The mid2008-2009 global economy crisis has triggered enormous interest in understanding the interactions between real and financial markets. Many authors use or extend one of canonical workhorse models with financial frictions to analyze how a presence of credit market imperfections can affect the real economic economy. E.g. a significant role of financial frictions in an economy can amplify responses of real variables to various shocks. Also during financial turbulence, shocks that initially hit financial variables can significantly propagate into real sectors and affect a position of an economy in business cycles.

In our contribution, we present a model with financial frictions and analyze it on Czech data. The model incorporates several features that are important for capturing Czech stylized facts along the balanced growth path. The most important are the real exchange rate appreciation in consumer prices, high openness, or gradual exchange rate pass-through modeled via a cascade of nominal rigidities and local currency pricing ([2]). Such framework is sufficiently rich to capture Czech data and its structural parameters are *structural* ([8]). The balance sheet channel is incorporated into the model through the financial accelerator mechanism. The debt-contracts between entrepreneurs and the financial intermediary follow the state non-contingent assumption ([3],[4]). The model also contains 'non-standard' shocks which should capture disturbances coming from the financial sector. Empirical analysis is carried out via analyzing some results from the Bayesian estimation. First, we discuss the choice of observables for the estimation. Then, we present Bayesian impulse responses to two shocks of the financial sector - a capital price bubble shock and higher riskiness of the entrepreneurial sector ('sigma' shock), and show estimation of both shocks.

2 Model description

This section presents the structure of the model. Since the model has relatively rich structure, we present only the financial part of the framework in greater detail. The model is developed for the inflation targeting regime with the conventional monetary policy. It has a balanced growth path where

[1]Corresponding author. Czech National Bank, Macroeconomic Forecasting Division, Na Příkopě 28, 115 03 Praha 1 and Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, e-mail: jiri.polansky@cnb.cz

[2]Czech National Bank, Macroeconomic Forecasting Division, Na Příkopě 28, 115 03 Praha 1 and Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, e-mail: jaromir.tonner@cnb.cz

[3]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, e-mail: osvald@econ.muni.cz

all variables are either constant or growth at some growth rate. To be consistent with Czech stylized facts, it incorporates the following frictions: habit formation in consumption, investment adjustment costs, gradual exchange rate pass-through which is modeled via a cascade of nominal rigidities ([2]) and the local currency pricing. The balance sheet channel is incorporated into the model through the financial accelerator mechanism. The debt-contracts between entrepreneurs and the financial intermediary follow the state non-contingent assumption ([3],[4]). The design of the model is shown in figure 1.

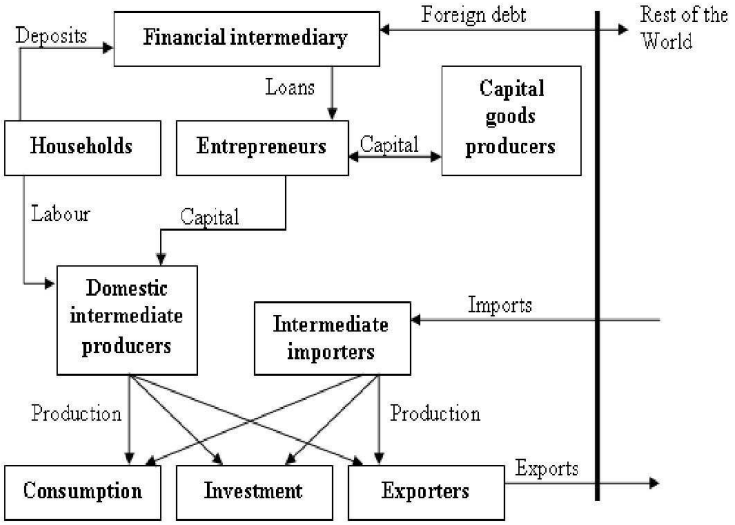


Figure 1: Model structure

Risk-averse households consume all varieties of consumption goods, choose the volume of deposits at the financial intermediary, supply a differentiated labour to domestic intermediate firms and set their nominal wages subject to the downward-sloping demand curve and quadratic adjustment costs. Also, households own all firms in the model. For the model to be tractable, households (and entrepreneurs) cannot store capital in time. Thus, the model contains perfectly competitive capital goods producers who own capital goods and sell it to entrepreneurs in each period. Investment decisions are made by households.

Entrepreneurs purchase physical capital ($P_t^K K_t$) from capital goods producers combining internal (net worth E_t) and external (bank lending L_t) funds

$$L_t = P_t^K K_t - E_t. \tag{1}$$

They have to use both sources of funding in each time as their net worth is not sufficient and cannot be sufficiently accumulated. Entrepreneurs rent the purchased capital to domestic intermediate firms for production purposes. Return to capital of each entrepreneur is affected by the idiosyncratic productivity ω whose value divides entrepreneurs into defaulting (insufficient return) and surviving. Thus, there is a cuttoff value

$$\omega^* \equiv \frac{R_{t+1}^L L_t}{R_{t+1}^K P_t^K K_t}. \tag{2}$$

where R_{t+1}^L is the lending rate and R_{t+1}^K is the return to capital.

Debt contracts between entrepreneurs and the financial intermediary is captured by the costly state verification ([9]) which determines the optimal behaviour between entrepreneurs who maximize expected profits and the financial intermediary who receives opportunity costs in expectation. This optimization problem incorporates the assumption of state non-contingent contracts ([3],[4]) under which the lending rate is fixed ex ante and cannot be changed after the aggregate return to capital is observed. Thus, the financial intermediary also bears risk of the contracts and makes profits or looses.[1] Under state non-contingent contracts, the maximization problem of entrepreneurs is subjected to a single constraint of the financial intermediary

$$\max_{K_t,\omega^*} \quad R_{t+1}^K P_t^K K_t[1-\Gamma(\omega^*)] + \lambda_t^C[R_{t+1}^K P_t^K K_t[\Gamma(\omega^*) - \mu G(\omega^*)] - R_t(P_t^K K_t - E_t)]$$

[1]The model does not contain capital of the financial intermediary.

where λ_t^C is the lagrange multiplier and μ are monitoring costs. The expected gross share of profits going to the financial intermediary is

$$\Gamma(\omega^*) \equiv \int_0^{\omega^*} \omega f(\omega) d\omega + \omega^* \int_{\omega^*}^{\infty} f(\omega) d\omega.$$

The net share of profits going to the financial intermediary is

$$\Gamma(\omega^*) - \mu G(\omega^*),$$

where the expected monitoring costs are

$$\mu G(\omega^*) \equiv \mu \int_0^{\omega^*} \omega f(\omega) d\omega.$$

The production structure of the economy contains domestic and import intermediate goods producing firms and three (consumption, investment, export) final goods producing firms. All sectors are monopolistically competitive. Such a cascade of nominal rigidities creates desirable interactions among production sectors and delivers multiple stages of exchange rate pass-through [2]. The domestic intermediate goods producers combine labour and capital via the Cobb-Douglas technology. The intermediate importers differentiate costlessly a single foreign good. The packed intermediate goods are purchased by final goods producing firms who utilize them as inputs for final production of consumption, investment and export goods. Following [2], the model incorporates the export-specific and openness technologies. The former captures the Harrod-Balassa-Samuelson effect which implies the real exchange rate appreciation in consumer prices in the steady-state. The latter captures an increase of the trade openness (trend in the nominal trade share in output).

Besides 'standard shocks', two shocks are incorporated into the model to deal with financial frictions issues. First, higher riskiness of the entrepreneurs is implemented by the stationary process for the standard deviation of the idiosyncratic productivity ω distribution

$$\sigma_t = \rho_\sigma \sigma_{t-1} + (1 - \rho_\sigma)\sigma + \varepsilon_t^\sigma.$$

Second, the model contains irrational capital price bubbles A_t^B modeled as exogenous deviations from the arbitrage condition ([5],[3])

$$P_t^{K_{obs}} = P_t^K \exp(A_t^B),$$

where $A_t^B = \rho_{A^B} A_{t-1}^B + \varepsilon_t^{A^B}$.

3 Empirical Analysis

3.1 Data

There are 16 time series which are used as observables for the empirical analysis. 3M PRIBOR and 3M EURIBOR are used for domestic and foreign interest rates. CZK/EUR is used as the observable for the exchange rate. Consumption, investment, exports, imports and price deflators for investment, export and import come from the Czech statistical office (CZSO). As the model does not contain government sector, the observable for private consumption contains consumption of households, expenditures of non-profit institutions serving households and also government consumption. Gross capital formation (GCF) is used as the observable for investment expenditures. It is a sum of gross fixed capital formation and a change of inventories. Within Czech national accounts data, GCF consists of private as well as government investment expenditures. To simplify the model, it is assumed that the consumption deflator level equals the headline CPI inflation level at each time. Thus, the headline CPI (CZSO) is used as the observable for both time series. The (seasonally adjusted) average nominal wage in the business sector is used as the observable for nominal wages. The effective euro area PPI (Consensus Forecast) is used as the observable for foreign inflation. Similarly, the effective euro area GDP (Consensus Forecast) is used for the calculation of the foreign demand observable.[2]

[2]The weights used in the calculation of the effective variables are the shares of the individual euro area economies in the foreign trade turnover of the Czech Republic [7].

To describe the behaviour of the financial sector, we choose the volume of loans and the lending rate as observables as the methodology of these time series is clear. In the ARAD database, these data are available from January 2004. Except for the short length, these data are relatively volatile and contain methodical changes. In our contribution, we assume that the most important firms in the economy can probably get loans even during the crisis and that these loans are big enough to influence the profile of the time series. Moreover, as banks restrict lending during the crises, loans to most important corporations are significant for them not only because of "a relatively safe return", but banks also try to keep those firms as their clients for future lending. Thus, we choose time series of loans up to 30 mil. and the corresponding lending rate as observables to link the financial accelerator with data. In 2010Q1, there is a methodical change in data of loans as revolving loans were taken away. Thus, the profile is adjusted by the expert judgment as otherwise there would be a problem with a model-consistent filtration.

3.2 Bayesian Estimation

The Bayesian estimation of the model parameters is carried out via the Dynare Toolbox [1]. The model has 44 structural parameters from which 33 is estimated. The rest of parameters is calibrated with respect to several reasons, e.g. estimation difficulties (discount factor), definition of the steady-state in the model (production shares in each final-goods producing sector), model behaviour (debt-elastic premium parameter) or micro data.

Table 1 shows model parameters associated with financial friction. The steady-state parameter of the capital to equity ratio $\frac{P^K K}{E}$ is calibrated with respect to Czech micro data. We do not use any observables of the capital stock, equity, or rate of defaults of entrepreneurs for the estimation. The calibration of the parameter for the steady-state of σ is set with respect to model properties (impulse responses) and a model-consistent filtration. Similarly, we do not estimate the posterior. Following [6], the prior for the monitoring costs parameter is set to 0.2 and the estimated value is slightly lower. The autoregressive parameters for both financial shocks (ρ_σ and ρ_{A^B}) indicate a relatively significant persistence.

	Parameter	Prior	Post. mean	Conf. int.	Distribution
$\frac{P^K K}{E}$	Capital to equity	1.8			
σ	Std. dev. of idios. product. dist.	0.45			
μ	Monitoring costs	0.20	0.1912	0.1752 - 0.2072	Gamma
ρ_σ	Persistence of σ	0.5	0.7719	0.5469 - 0.9592	Beta
ρ_{A^B}	Capital price bubble persistence	0.5	0.5403	0.2930 - 0.7833	Beta

Table 1: Financial frictions parameters

Figure 2 shows Bayesian impulse responses to the capital price bubble shock. Higher observed price of capital implies an increase of investment expenditures and a decrease of private consumption. Investment expenditures increase and private consumption decreases. Nominal exchange rate depreciates and the monetary authority increases interest rates to bring the headline inflation back to the target.Nominal wage growth is temporarily lower.

Figure 3 shows impulse responses to a σ shock which approximates higher riskiness of the entrepreneurial sector. Investment expenditures significantly decrease. As investment is produced mostly from imports, net exports increase. Exchange rate appreciates and the monetary authority decreases interest rate to bring inflation back to the target. Consumption expenditures of households slightly increase. The reaction of a majority of variables is much lower than the response of investment as financial frictions shocks affect primarily investment expenditures.

Figure 4 shows profiles of shocks on Czech data. During the global economy crisis, the smoother identifies a significant riskiness shock in 2009Q1 and moderate shocks during the rest of 2009. During the recovery of investment expenditures in 2010, there is a significant negative shock. The capital price bubble shocks are estimated most significant in the 2008-2009 turn with the peak in 2008Q3.

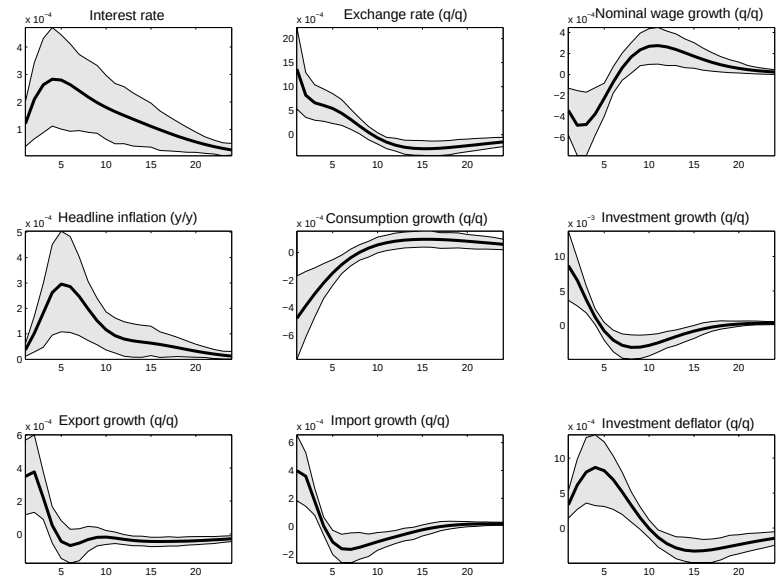


Figure 2: Capital price bubble shock

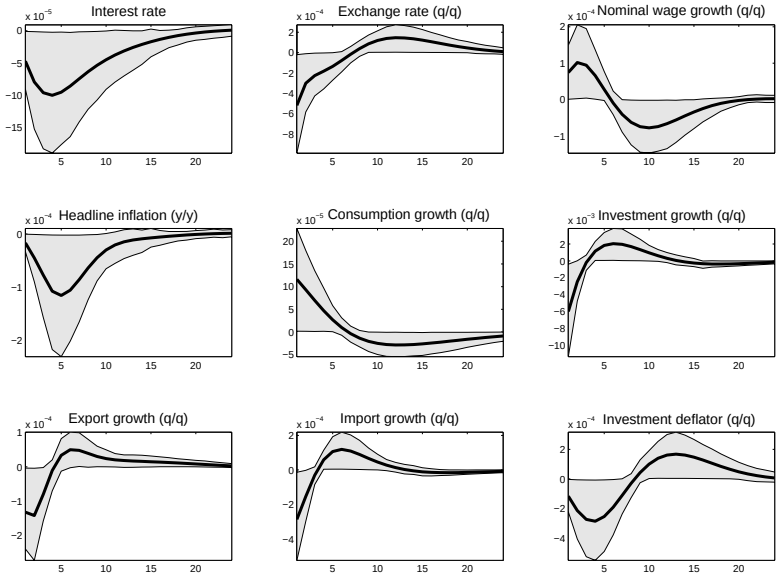


Figure 3: Sigma shock

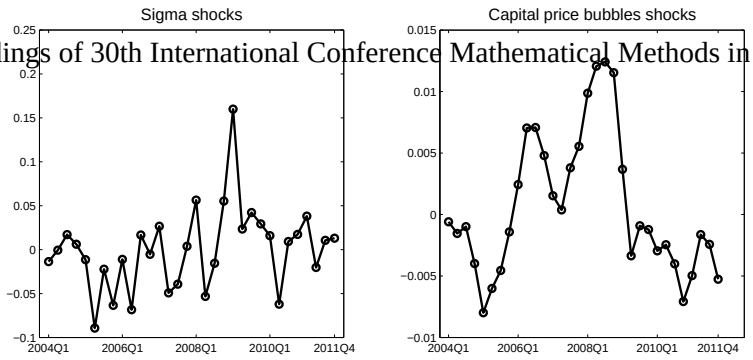


Figure 4: Sigma and capital price bubble shocks

4 Conclusion

In our contribution, we present structure of the model with the financial accelerator which is tailor-made for the czech economy. Besides several features capturing the main trend-cyclical components of Czech stylized facts, it contains state non-contingent debt-contracts between entrepreneurs and the financial intermediary. We estimate the model on Czech and euro area data and show Bayesian impulse responses to two shocks associated with financial frictions.

Acknowledgements

The financial support of specific research at Faculty of Economics and Administration, project MUNI/A/0780/2011 and the Research Project B2/10 at the Czech National Bank are gratefully acknowledged. We are grateful to Mirek Beneš and Jan Brůha for many helpful comments and advices.

References

[1] Adjemian, S., Bastani, H., Juillard, M., Mihoubi, F., Perendia, G., Ratto, M., and Villemot, S.: Dynare Reference Manual, Version 4, CEPREMAT, Dynare Working Papers, 2011.

[2] Andrle, M., Hlédik, T., Kameník, O., and Vlček, J.: Implementing the New Structural Model of the Czech National Bank. Czech National Bank, Working Paper Series 2, 2009.

[3] Beneš, J.: Macro-Financial Linkages and Macro-Prudential Policy. Workshop in the Czech National Bank, 2010.

[4] Beneš, J. and Kumhof, M.: Risky Bank Lending and Optimal Capital Adequacy Regulation. IMF Working Paper 11/130, 2011.

[5] Bernanke, B. S. and Gertler, M.: Monetary Policy and Asset Price Volatility, Federal Reserve Bank of Kansas, 1999.

[6] Bernanke, B. S., Gertler, M., and Gilchrist, S.: The Financial Accelerator in a Quantitative Business Cycle Framework. In: *Handbook of Macroeconomics* (J. B. Taylor and M. Woodford, eds.), North Holland, 1999, 1341–1393.

[7] Czech National Bank: Inflation Reports.

[8] Tonner, J., Polanský, J., and Vašíček, O.: Parameter Drifting in a DSGE Model Estimated on Czech Data, *Czech Journal of Economics and Finance* **61** (2011), no. 5, 510–524.

[9] Townsend, R. M.: Optimal Contracts and Competitive Markets with Costly State Verification, *Journal of Economic Theory* **21** (1979), 265–293.

Comparison of different non-statistical classification methods

Ondřej Popelka[1], Jiří Hřebíček[2], Michael Štencl, Michal Hodinka, Oldřich Trenz

Abstract. In this article, we aim to compare different methods usable for solving classification problems. A substantial number of methods that are not based on mathematical statistics may be used. Exploring these methods is interesting, because they are often capable of solving problems, which are not easily solvable using classificators based purely on mathematical statistics.
There are many approaches available such as support vector machines, neural networks, evolutionary algorithms, parallel coordinates, etc. In this article, we concentrate on describing different neural network approaches, parallel coordinates and genetic algorithms. Neural networks come in many flavors (e.g. multi-layer perceptron, non-linear autoregressive networks) and they have achieved some recognition. Genetic algorithms also have been used for classification many times before, but with mixed results. In this article, we describe and evaluate different capabilities of these methods when used for economic data. This for example includes identification of hidden data structures, dealing with outliers and noise.

Keywords: classification, decision trees, neural networks, parallel coordinates, corporate performance, sustainability reporting.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

One of our current research projects is project No P403/11/1103 "Construction of Methods for Multi-factorial Assessment of Company Complex Performance in Selected Sectors" solved by Faculty of Business and Economics (FBE) at Mendel University in Brno in cooperation with Faculty of Business and Management (FBM) of Brno University of Technology (BUT). The project is funded by Czech Science Foundation and solved during the years 2011–2014. There are six main research targets defined in [14], one of which is the construction of quantitative and qualitative methods of the multifactor measurement of corporate performance.

To achieve the stated research goal we have analyzed corporate performance factors in chosen companies in the real estate and construction sectors. Only companies that have successfully implemented international management standards [13] such as ISO 9000 (quality), ISO 14000 and EMAS (environmental), ISO 18000 (health and safety) were evaluated. The performance factors including Environmental, Social and Governance (ESG) factors [18], [5] are being transformed to *Key Performance Indicators* (KPIs) [11], [15], which are organized into different standardized categories (economic, environmental, social, etc.).

One of the tasks, which are part of the development of multifactor measurements of company performance is solving of a classification problem on a multivariate dataset. In this article, we describe several methods usable for classification and our experiments with these methods. We focus on data-mining methods, because our primary sources of data are company questionnaires and reports. Therefore, our data may contain errors, missing values and other features, which may present [24] a difficulty for processing by pure statistical methods such as k-means.

2 Experiments

In our experiments, we decided to include these methods: neural networks, decision trees, support vector machines and genetic algorithms. This choice was motivated by some preliminary research [24] and by the methods used by the authors of the test dataset [19]. We have used a testing dataset for two reasons – firstly because we aim to conduct an independent comparison and secondly because our data collection is still in progress. In [19] the authors used the Support vector machine (SVM) method, Naive Bayes and a Decision Tree

[1] Department of informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, e-mail: ondrej.popelka@mendelu.cz, jiri.hrebicek@mendelu.cz, michael.stencl@mendelu.cz, michal.hodinka@mendelu.cz, oldrich.trenz@mendelu.cz.

to analyze the presented data. Software used was *rminer* library for R tool [7]. For our experiment we have used standard implementations in Weka data mining software suite for all algorithms apart from Grammatical Evolution (chapter 2.5), because that has no standardized implementation. The SVM algorithm is implemented using the LibSVM library.

2.1 Source data

As mentioned above, the dataset used for experiments in this paper, was collected by the authors of [19]. It contains results of direct bank marketing campaigns. Specifically it includes 17 campaigns of a Portuguese bank conducted between May 2008 and November 2010. These were always conducted over the phone and the customer was offered a long-term deposit application. Originally the dataset contained 79 354 contacts, out of which contacts with missing data or inconclusive results were discarded leading to a dataset with 45 211 instances with two possible outcomes – either the client signed for the long-term deposit or not.

The dataset contains 16 input variables. There are eight variables related to the client:

- age (numeric),
- job type (categorical),
- marital status (categorical),
- education (categorical),
- whether the client has credit in default (binary),
- average yearly balance in Euros (numeric),
- whether the client has a loan (binary),
- whether the client has a personal loan (binary).

Four variables relate to the last contact of the current campaign:

- contact communication type (categorical),
- last contact day of the month (numeric),
- last contact month of year (categorical),
- last contact duration in seconds (numeric).

There are four more variables regarding the campaign:

- number of contacts performed during this campaign and for this client (numeric),
- number of days that passed by after the client was last contacted from a previous campaign (numeric),
- number of contacts performed before this campaign and for this client (numeric),
- outcome of the previous marketing campaign (categorical) [19].

As mentioned above, the output variable corresponds to campaign output, which has been reduced to a binary output. For testing of the classification algorithms, the dataset was split into 2/3 of training data and 1/3 of test data. This leaves us with 31 489 randomly chosen instances for training and 13 722 instances for testing. Baseline classifier (ZeroR classifier) which selects all instances into the largest class has an accuracy of 88.30%. This is caused by strong skew of the output variable. It corresponds to the fact than only 11.70% of bank clients agreed to sign for the long-term deposit (in the unfiltered data, the success rate is 8%) [19]. It is important to note that such strong skew in the dataset provides a considerable challenge for the classification algorithm.

2.2 Decision trees

As a base starting method, the decision trees were used, because this is a very widely used classification technique. A decision tree algorithm is a technique which recursively splits the dataset instances into classes until all the data are assigned to a class. Basic approaches are breadth-first or depth-first greedy searches. Several algorithms may be used to construct a decision tree. The most common method [25] is the C4.5 algorithm [22][25], which is an improved version of the older ID3 algorithm. Furthermore, there are a number of algorithms, which claim to have better accuracy, such as ADTree [10] or Random Forests [3]. An in-depth comparison of several decision tree algorithms can be found in [2] and [22]. For the sake of comparison, we have also tested a BFTree algorithm that implements best-first decision tree classifier.

An Alternating decision tree algorithm (ADTree) was designed especially so that the resulting tree is intelligible to humans [10]. It is a combination of a weak classifier based on decision trees and decision stumps with a boosting learning algorithm (ADTBoost). This is done by replacing an ordinary decision tree with an alternating tree that uses prediction sums sign instead of the class label itself as a leaf node. Therefore, each instance is mapped to a real value prediction, which is a sum of the predictions of the base rules in the instance set. The actual label is replaced by the sign of the sum of the predictions. This transformation has two main

effects. Without an extension, the ADTree supports only two-class problems (binary class). Each rule has assigned a confidence value – classification margin – that suggests the reliability of the rule.

2.3 Support Vector Machine

Support vector machine is a learning algorithm originally designed only for linear classification models. The algorithm tries to find objects (support vectors) which define hyper-planes separating a set of any two classes with maximum margin. Nonlinear classification problems may be solved by transforming the search space into a transformed feature space. This transformation is done using the kernel basis function, where several options are available [16]. Several options are available for the partitioning kernel function – linear function, polynomial function with different degrees, B-spline function and radial basis function (RBF) [16]. However, during our experiments none of the mentioned transformations led to a classification better than the one of the baseline classifier. This is further discussed in the Results section.

2.4 Neural networks

Neural networks are one of the best-known representatives of learning algorithms and as such, they may be divided into two basic classes – unsupervised and supervised networks. The former are commonly represented by Competitive layers or Self-organizing maps which automatically find relationships within the input vectors [12]. Given the fact that the processed data are labeled and we have a priori knowledge of the output classes, we have concentrated on supervised networks, which exhibit better performance in such cases [9]. There are many flavors of supervised networks, where by far the most common one is still Multilayer Perceptron (MLP) networks with back-propagation learning algorithm. Additionally in our previous research [23], we have been testing other approaches such as Radial Basis Function (RBF) and Nonlinear autoregressive neural network (NAR).

In [23][23], it has been demonstrated, that RBF neural networks exhibit similar performance to MLP networks while having a substantially better performance. RBF networks are feed-forward networks in which perceptrons are replaced by local units. These local units use radial functions which, given a central point, provide the same output for arguments with the same distance from the central point. RBF networks generally have only one hidden layer that makes the back-propagation learning algorithm much simpler and faster.

2.5 Genetic algorithms

Genetic algorithms represent a broad area of methods and algorithms. Genetic algorithms are mostly used as optimization algorithms. However, they may also be used for classification. There have been many experiments performed in this area such as [6]. However, genetic algorithms have suffered a noticeable decline in popularity. More precisely rather than being used as a sole classification algorithm, they are used as part of hybrid methods, for example in: [17], [4] and [1].

We have however decided to use a slightly more specific flavor of genetic algorithms – grammatical evolution. This algorithm uses a genetic algorithm to control a generative context-free grammar. That is, a context-free grammar is defined, which may be used to generate strings in an arbitrary format, for example “if (condition and condition) then class1 else class2”. This is defined using a set of terminals, non-terminals and production rules that define the language in which the strings may be generated. A genetic algorithm is then used to generate various strings of the language, which are evaluated and assigned a fitness value (in this case, classification accuracy was used).

The algorithm is extensively described in [20]. Further, it is extended to generate numeric constants by using two-level grammatical evolution as described in [21]. This is important when solving this classification problem, because the conditions are to be in format “if (age > 20) then will-sign else won’t sign”. That is, each condition expression consists of a variable (see chapter 2.1) and a constant value of that variable which defines class border. A limitation in the current implementation requires that the constant values are numeric, therefore for this experiment, the text attributes were converted to discrete numeric variables.

As stated above, the algorithm output is arbitrary, so it may be same as a decision tree model or it may be generated so that it is directly usable in a common programming language (C++, Java, PHP, etc.). On the other hand, this means that the task is much more difficult. A terminal criterion for the genetic algorithm was set to either classification accuracy of 89.2% or computation time of 8000 seconds. We have set the time criteria arbitrarily to an acceptable time of computation (which is still about 1000 times bigger than computation of ADTree). All of the ten runs performed on GE algorithm stopped on the time criterion and have not reached classification accuracy better than the baseline classifier. This is a bit of a disappointment, but not surprising, Because GE is a very generic string generation algorithm not optimized for classification problems.

3 Results

In our experiments, we have investigated several methods for automatically classifying a multivariate dataset. The results are summarized in Table 1 that shows different tested methods. For each method the *classification accuracy* (amount of correctly classified instances), *the time required* to finish the classification and *ROC area* (area under the ROC curve) is shown.

Algorithm	Classification accuracy	Time required [s]	ROC Area
ADTree	89.57%	6.6	0.889
RandomForest	89.57%	3.5	0.892
BFTree	89.61%	100.1	0.730
C4.5 (J48)	90.20%	2.7	0.839
RBF	89.27%	4.0	0.826
MLP	89.26%	706.8	0.877
SVM	88.23%	2618.6	0.500
GE	88.23%	> 8000	0.500

Table 1: Results of comparison, average over 10 runs for best performing configuration.

In our comparison, we have focused on methods that can provide the information about the exact classification rules. The method should provide at least the information about more significant and less significant for the classification. This kind of information is generally best obtained from decision trees that provide classification rules in a format easy to visualize and understand.

Experiments with decision trees can be considered highly successful. We have obtained slightly better results than the authors of the data in [19] had, who achieved ROC Area of 0.868. As mentioned above, we are also interested in providing the classification rules in readable format. For this, it is interesting to compare the sizes of decision trees, which are shown in Table 2. There is a strong disproportion of the size of the decision tree across multiple algorithms. The Random Forest algorithm is not shown in the results, because it consists of several trees. However, the total number of nodes was higher than 10 000. An easy to understand set of rules is obtained from the Alternating decision tree (ADTree). From this set, it is clear that important variables with high discriminative power are: *duration of the call*, *month of contact*, *previous outcome*, *whether the customer has a housing loan* and *contact type*. This is consistent with the original results.

Algorithm	Number of nodes	Number of leaves
ADTree	31	21
BFTree	279	140
C4.5	1168	1716

Table 2: The size of decision trees for different algorithms.

The SVM based classifier proved the least successful as it did classify the same as baseline classifier. This is a considerable surprise, since in the original paper [19] the SVM based classifier was the best performing (although on a smaller dataset). Overcoming the high memory consumption problem of the SVM classifier was enabled by using 64-bit architecture. A possible explanation of the poor classification performance is that in [19] there were 29 input variables used. In our experiments, we have used only 16 publicly available input variables. We will investigate this problem in our future work. Because the classification was not successful, we did not attempt to extract the decision rules, although this is possible using [8].

The results of neural networks are perfectly comparable with those of decision trees in terms of both precision and performance. However, the RBF network greatly outperforms the MLP type of network, which is consistent with the results of [23]. Unfortunately, a big trade of RBF networks is that it does not provide any information about the classification rules. The MLP network does not provide the rules either, but still some information may be extracted from the weights assigned to different variables. From this, we can infer, that *duration of the call*, *month of contact* and *previous outcome* or *housing loan* are variables consistently with above average weights. Again, this result is consistent with both decision trees and original results.

The importance or discriminative power of variables may also be inferred from preliminary data analysis. We used the principles of exploratory data analysis (EDA) and visualizations such as parallel coordinates and Polyviz visualizations. Although without additional software, the importance of the variables cannot be quantized. Still these tools provide an invaluable insight into the dataset, which is useful especially for visual data mining and further to verify the results of the decision tree algorithms.

The conclusion can be stated that the favorite method so far is Alternating decision tree because with computational time under 10 seconds (for 45 211 instances) it belongs to the group of algorithms with very good performance. In addition, the classification performance is the second best of the tested algorithms. Furthermore, the classification rules are provided in easy to understand format and their amount is completely acceptable to a human being. This is a very important property of the algorithm, because in our task of constructing multifactor measurements of company performance we need to ensure that such performance measurements are presentable to the company management. It may come as a surprise that the more sophisticated methods were outperformed by simpler methods; nevertheless, we still consider it an important result.

Acknowledgements

This paper is supported by the Czech Science Foundation. Name of the Project: Construction of Methods for Multifactor Assessment of Company Complex Performance in Selected Sectors. Registration No: P403/11/2085.

References

[1] Ahna, H., and Kim, K.: Bankruptcy prediction modeling with hybrid case-based reasoning and genetic algorithms approach, *Applied Soft Computing* Volume 9, Issue 2 (March 2009), 599–607.

[2] Anyanwu, M. N., and Shiva, S.: Comparative Analysis of Serial Decision Tree Classification Algorithms, *International Journal of Computer Science and Security (IJCSS)* Volume 3 Issue 3 (2009), ISSN 1985-1553, 230–240.

[3] Breiman, L.: Random Forests, *Machine Learning* Volume 45 Issue 1 (2001), Springer, 5–32.

[4] Chen, M. Classification Techniques of Neural Networks Using Improved Genetic Algorithms. In Proceedings of WGEC '08. Second International Conference on Genetic and Evolutionary Computing, 2008, ISBN 978-0-7695-3334-6, 115–119.

[5] Chvátalová, Z., Kocmanová, A., and Dočekalová, M.: Corporate Sustainability Reporting and Measuring Corporate Performance. In: *Proceedings of Environmental Software Systems. Frameworks of eEnvironment. 9th IFIP WG 5.11 International Symposium. ISESS 2011*, Springer, Heidelberg, 2011, 398–406.

[6] Corcoran, A.L.: Using real-valued genetic algorithms to evolve rule sets for classification. In: *Proceedings of Evolutionary Computation - IEEE World Congress on Computational Intelligence*, Orlando, USA. 1994, ISBN 0-7803-1899-4, 120–124.

[7] Cortez, P.: Data Mining with Neural Networks and Support Vector Machines using the R/rminer Tool. In: *Proceedings of the 10th Industrial Conference on Data Mining*, Berlin, Germany, July 2010, Springer, LNAI 6171, 572–583.

[8] Cortez, P., and Embrechts, M.: Opening Black Box Data Mining Models Using Sensitivity Analysis. In: *Proceedings of IEEE Symposium on Computational Intelligence and Data Mining*, pp. 341–348. Paris, France, 2011.

[9] Fejfar, J., Šťastný, J. Cepl, M. Time series classification using k-Nearest Neighbours, Multilayer Perceptron and Learning Vector Quantization algorithms. In: *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* Volume 60 Issue 2 (2012), ISSN 1211-8516, 69–72.

[10] Freund, Y., Mason, L.: The alternating decision tree learning algorithm. In: *Proceeding of the Sixteenth International Conference on Machine Learning*, Bled, Slovenia, 1999, 124–133.

[11] Garz, H., Schnella, F., and Frank, R.: KPIs for ESG. *A Guideline for the Integration of ESG into Financial Analysis and Corporate Validation*. Version 3.0, Frankfurt, DVFA/EFFAS, [Online], Available: http://www.dvfa.de/files/die_dvfa/kommissionen/non_financials/application/pdf/KPIs_ESG_FINAL.pdf, 2010.

[12] Haykin, S. O.: *Neural Networks and Learning Machines*, 3rd edition, Prentice Hall, ISBN 9780131471399, 2009.

[13] Hřebíček, J., Soukopová, J., and Kutová, E.: Standardization of Key Performance Indicators for Environmental Management and Reporting in the Czech Republic. *International Journal of Energy and Environment* Volume 4 Issue 4 (2010), 169–176.

[14] Hřebíček, J., Soukopová, J., Štencl, M., and Trenz, O.: Corporate Performance Evaluation and Reporting. In: *Proceedings International Conference on Environment, Economics, Energy, Devices, Systems, Communications, Computers, Pure and Applied Mathematics* (*WSEAS)*, Wisconsin, USA, 2011, 338–343.

[15] Hřebíček, J., Soukopová, J., Štencl, M., and Trenz, O.: "Corporate Key Performance Indicators for Environmental Management and Reporting," *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* Volume 59 (February 2011), 99–108.
[16] Ivanciuc, O.: Applications of Support Vector Machines in Chemistry. In: *Reviews in Computational Chemistry* Volume 23 (2007), Wiley-VCH, Weinheim. 291–400.
[17] Kalia, H., Dehuri, S., and Ghosh, A.: Multi-Objective Genetic Algorithms for Fuzzy Classification Rule Mining: A Survey. *The IUP Journal of Information Technology* Volume 7 Issue 1 (March 2011), 7–34.
[18] Kocmanová, A., Dočekalová, M., Němeček, P., and Šimberová, I.: Sustainability: Environmental, Social and Corporate Governance Performance in Czech SMEs. In: *Proceedings of The 15th World Multi-Conference on Systemics, Cybernetics and Informatics* (WMSCI 2011), Orlando, USA, 2011, 94–99.
[19] Moro, S., Laureano, R., and Cortez, P.: Using data mining for bank direct marketing : an application of the CRISP-DM methodology. In: *Proceedings of the European Simulation and Modelling Conference – ESM'2011 (*EUROSIS), Guimarães, Portugal, October, 2011, 117–121,
[20] Popelka, O., and Ošmera, P.: Parallel Grammatical Evolution for Circuit Optimization. *Lecture Notes in Computer Science* Volume06 Issue 5216 (2008), ISSN 0302-9743. 425–430.
[21] Popelka, O. Two-level optimization using parallel grammatical evolution and differential evolution. In: *Proceedings of Mendel '07, 13th International Conference on Soft Computing*, Brno, 2007, ISBN 978-80-214-3473-8, 88-92.
[22] Quinlan, R.: C4.5: *Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA. 1993.
[23] Štencl, M., Popelka, O., and Šťastný, J.: Comparison of time series forecasting with artificial neural network and statistical approach. In *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* Volume 59 Issue 2 (2011), ISSN 1211-8516, 347-352.
[24] Štencl, M., and Šťastný, J.: Artificial Neural Networks Numerical Forecasting of Economic Time Series. *Artificial Neural Networks – Application. Artificial Neural Network*. Riejka, Croatia, InTech, 2011. ISBN 978-953-307-188-6. pp. 13–28.
[25] Venkatadri, M., and Lokanatha, C. R.: A Comparative Study On Decision Tree Classification Algorithms In Data Mining. *International Journal Of Computer Applications In Engineering, Technology And Sciences (IJ-CA-ETS)* Volume 2 Issue 2 (2010), ISSN 0974-3596. 24–29.

Usage of the extremal algebra in solving the travelling salesman problem

Alena Pozdílková[1], Richard Cimler[2]

Abstract. This article compares many ways of solving the traveling salesman problem. At first classical heuristic methods and methods using graph theory are mentioned. At the first part many universal methods are described, which can be also used in other transportation problems. The traveling salesman problem is solved by using genetic algorithm in the second part of this article. This algorithm generates at the beginning the first generation, chooses five thousand parents by the roulette method, crosses these pairs and determines the next generation. This process continues with next generations until stabilization. The algorithm is demonstrated on two examples. At the final part extremal algebras, max-plus algebra and max-min algebra, are defined and illustrated by examples. Monge matrices and some their properties are described in these algebras. The optimization of the traveling salesman problem, which leads to a reduction of the computation complexity, is described for these matrices. The optimization is solved at first for the classical case and then for the matrices that satisfy the strict Monge property. Matrices with the strict Monge property lead to the linear complexity of the problem.

Keywords: Monge matrix, max-plus algebra, max-min algebra, travelling salesman problem, genetic algorithm.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Different ways of solving the travelling salesman problem are compared in this paper. Classical heuristic algorithms, genetic algorithms and usage of Monge matrix in extremal algebras are described.

2 The travelling salesman problem

The travelling salesman problem can be in practice encountered very often. In comparison with realization each travel separately, the ring connection can save a great deal of costs in cases where exists the necessity to distribute a certain material from one or more producers to many consumers.

The travelling salesman problem represents the problem of finding the shortest ring connection among M cities where their distance is known in advance. Each distance in fact represents the cost of every route (the assumption is that the costs to travel from city A to city B are the same as vice versa). The goal is to minimalize the total costs.

The problem doesn't lie in discovering the algorithm of finding the shortest connection – the easiest approach is obvious: to search among all the circuits among the given cities and choose the shortest of them.

The problem lies in the fact that with the mounting number of cities the number of potential circuits increases very quickly. Thus the time needed to compute the aforesaid brutal force solution becomes quite unacceptable already with a couple of dozen cities.

Resources to this chapter are [1], [5], [6], [9]. The next chapter contains the listing of known methods to solve this problem.

2.1 Standard solving methods

The travelling salesman problem belongs to NP-hard tasks. These tasks can be easily formulated but they are

[1] University of Hradec Králové, Department of Information Technologies FIM, Rokitanského 62, 500 03 Hradec Králové, alena.pozdilkova@uhk.cz.
[2] University of Hradec Králové, Department of Information Technologies FIM, Rokitanského 62, 500 03 Hradec Králové, richard.cimler@uhk.cz.

difficult to solve. They can be solved using heuristic methods, general methods, methods based on the graph theory or by means of genetic algorithms.

2.2 Heuristic methods

The outcome of heuristic methods can be considered to be an acceptable solution, but the value of the objective function may not be optimal. This fact is their principal disadvantage. Contrary to exact methods, the heuristic methods are fast, polynomial and it is easy to use them to solve even the complex tasks. The main and essential disadvantage is the fact that they don't guarantee discovering the solution which is the global optimum. Another important disadvantage is the impossibility to determine the estimation of the error.

Heuristic methods can be split to two methods. First are the solutions composing methods which construct the solution from the beginning. The other ones are the solution improving methods which are based on a certain solution, which is then improved by iterative procedures. Solution composing methods can be further split to methods of the sequential procedures, which operate locally and take into consideration only the vicinity of the partial result and methods of the parallel procedures which operate globally – they start at several places simultaneously and then combine the partial solutions into the global one.

2.3 General methods

These methods can be commonly used even for different types of transport tasks provided that the costs of the travel from city A to city B are the same as the costs from city B to city A.

The disadvantage of these methods is the fact that there is not known any estimation of precision of the solution.

2.4 Methods based on the graph theory

The second big group of algorithms which is possible to use for solving the traveling salesman problem are methods based on the graph theory. Fastest are greedy algorithms which work very simple. At first the edge with minimal costs is chosen. This edge is added to result way and is removed from further computing.

2.5 Evolution methods

Next way how to solve the travelling salesman problem is usage of the evolution algorithms. It is suitable to use the natural representation. For example the way B - A - D - E – C is represented as a chromosome string (b a d e c). Usage of special selection operators is also advantageous. One of the most suitable is recombination operator with edge crossing - it is the most successful operator for solving the travelling salesman problem. Its specialty is that two parents create only one descendant. Each segment of route comes from one of the parents. For each city the edge table stores information about its neighbors. List of neighbors comes from its parents.

One of the cities with minimum number of neighbors is randomly is chosen. This city is added to result way and is deleted from the edge table and then algorithm repeats this procedure again.

Algorithm finds the minimal possible ring way between specified number of points. Roulette wheel is used for choosing parents. Scales of roulette wheel are inverse proportional to the length of the way. The ways with smaller length have bigger rating on the scales. For example half of the distance has double bigger rating.

Chosen pairs of parents are crossed by selection operators with the edge recombination. This crossing creates one new descendant. Only in the case that the length of its way is smaller than length of the way of his parents then descendant is part of the next generation otherwise parent with minimal length of the way is part of the next generation. This algorithm continues until the certain accuracy is reached. Our example shows that after few steps there is no further improvement of length of the way. This algorithm can lead only to discover of the local optimum instead of the global optimum.

3 Implementation of genetic algorithm

- First generation is randomly generated
- 5000 pairs of parents are chosen by roulette wheel method
- Crossing of parents and forwarding to next generation one of parents or descendent – depends on shortest way
- Repeating this algorithm until stable state is reached. Local optimum – shortest path is found.

This algorithm was programed in the Pearl and in the Matlab. The results of the experiments are shown at following part of this paper.

3.1 Results of the first experiment

10 cities (square net 1000 x 1000), 5000 parents, 5 generations (after that stable state has been reached)

Cities: A 744 299, B 88 294, C 663 996, D 981 654, E 509 554, F 387 92, G 547 410, H 806 168, I 420 186, J 227 864

- Generation 0 - J I F B E G A H D C = 3619.28216907124
- Generation 1 - F I B E J C D A H G = 3567.40119270441
- Generation 2 - J B F I E G A H D C = 3384.84611373796
- Generation 3 - A H I F B G E J C D = 3382.63430361154
- Generation 4 - F B J C D E G A H I = 3358.5860296155
- Generation 5 - F B J C D E G A H I = 3358.5860296155

As you can see in generation 4 the shortest route was found.

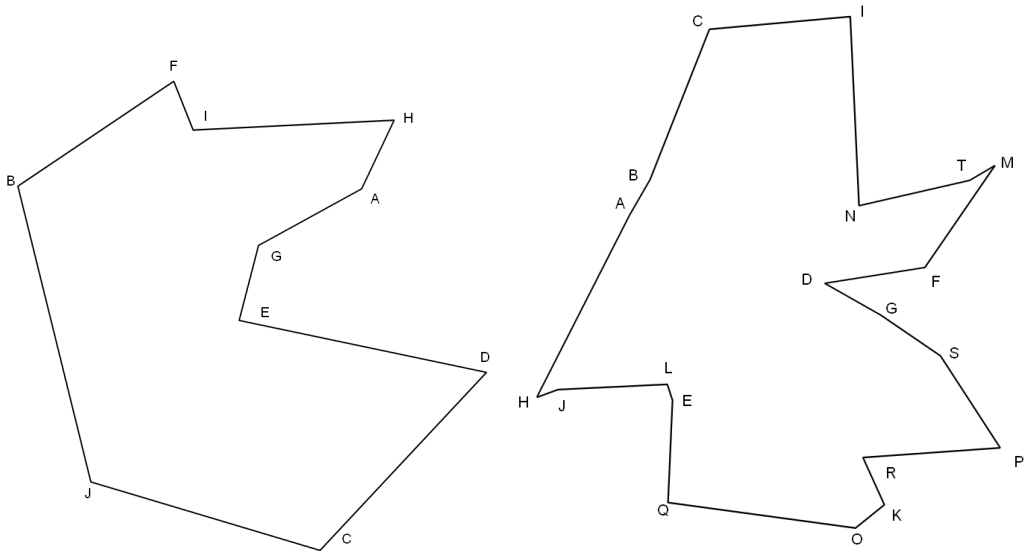
3.2 Results of the second experiment

20 cities (square net 1000 x 1000), 5000 parents, 20 generations (after that stable state has been reached)

Cities: A 200 350, B 237 288, C 340 29, D 537 468, E 275 668, F 709 441, G 635 522, H 42 663, I 583 7, J 77 651, K 639 851, L 266 642, M 832 263, N 595 333, O 590 890, P 841 752, Q 268 845, R 604 768, S 737 592, T 788 290

- Generation 0 - E A B C I Q J L H T R P G S D M F N K O = 6455.4108624135
- Generation 1 - Q P S N T G E B C I M F O K L R D A J H = 6041.12292335198
- Generation 2 - F M Q E L J H T N I C A B P D S G K O R = 5876.67183903036
- Generation 3 - P R S G T M D F O K N I C B A E Q L J H = 5149.89044555545
- Generation 4 - P R S G T M D F O K N I C B A E Q L J H = 5149.89044555545
- Generation 5 - F D R K O S T M P N Q E H J L A B C I G= 5064.88710405238
- Generation 6 - F M T N R G S I C B A J H Q E L K O P D = 4696.05677939314
- Generation 7 - S R P K O Q E J H L A B D G F M T N C I = 4363.16256524023
- Generation 8 - S G T M N F P D I C B A E L J H Q K O R = 4317.18528791843
- Generation 9 - S F T M N D G I C A B L E J H Q K O P R = 4152.42657949083
- Generation 10 - F S P O K D R Q E L J H A B C I N T M G = 4129.020628350
- Generation 11 - Q E L H J A B C I T M N F S P D G K O R = 3923.424919074
- Generation 12 - O K P R S D G F N T M I C B A L E Q J H = 3967.5052191970
- Generation 13 - O K P R S D G F N T M I C B A L E Q J H = 3967.5052191970
- Generation 14 - S G F N D T M I C B A J H E L Q K O P R = 3852.7588661828
- Generation 15 - S G F N D T M I C B A J H E L Q K O P R = 3852.7588661828
- Generation 16 - C B A H J L E Q O K R P S G D F M T N I = 3484.2585837226
- Generation 17 - C B A H J L E Q O K R P S G D F M T N I = 3484.2585837226
- Generation 18 - C B A H J L E Q O K R P S G D F M T N I = 3484.2585837226
- Generation 19 - C B A H J L E Q O K R P S G D F M T N I = 3484.2585837226
- Generation 20 - C B A H J L E Q O K R P S G D F M T N I = 3484.2585837226

The shortest route was found in generation 16.



Picture 1: Illustration of the first and the second experiment – the shortest route

In comparison with solutions using extremal algebras, this algorithm is very fast but it can find sometimes only the local optimum instead of the global optimum. By using the Monge matrices in extremal algebras, used in next chapter, we are always able to find the optimal solution in the polynomial time.

4 Extremal algebras

In the next part of this paper extremal algebras are described. Following text is based on information from [2], [3], [4], [8].

4.1 What is max-plus algebra

Max–plus algebra $(\bar{R},\oplus,\otimes)$ is an algebraic structure with two binary operations $\oplus$, $\otimes$ and a set $\bar{R} = R \cup \{-\infty, \infty\}$, which is an extension of real numbers. The operations maximum and plus in max-plus algebra are derived from the linear ordering in the set of real numbers. Max-plus algebra belongs to the family of so-called extremal algebras.

The operations in max-min algebra are defined as follows:

$$\text{for } x, y \in \bar{R}\text{: } x \oplus y = \max(x, y),\ x \otimes y = x + y.$$

For matrices A, B over $\bar{R}$ we can define operations $\oplus$, $\otimes$ analogously as in linear algebra over R with addition and multiplication. We assume matrices A, B of suitable types.

Example 1.

$$(4 \quad 8 \quad 2) \otimes \begin{pmatrix} 5 \\ 3 \\ 9 \end{pmatrix} = (4 \otimes 5) \oplus (8 \otimes 3) \oplus (2 \otimes 9) = 9 \oplus 11 \oplus 11 = 11$$

4.2 What is max-min algebra

Analogically we can define max-min algebra, where operations are defined as follows:

$$\text{for } x, y \in \bar{R}\text{: } x \oplus y = \max(x, y),\ x \otimes y = \min(x, y).$$

Example 2.

$$(4 \quad 8 \quad 2) \otimes \begin{pmatrix} 5 \\ 3 \\ 9 \end{pmatrix} = (4 \otimes 5) \oplus (8 \otimes 3) \oplus (2 \otimes 9) = 4 \oplus 3 \oplus 2 = 4$$

4.3 Monge matrices in max-min algebra

Monge matrices in max-min algebra are special matrices which hold following condition: for all elements $a \in A$ and $i, j, k, l \in \mathbb{N}$, where A is a matrix of type (m, n), i, k are row indexes, where $i < k$ and j, l are column indexes, where $j < l$:

$$a_{i,j} \otimes a_{k,l} \leq a_{i,l} \otimes a_{j,k}$$

This condition has to be fulfilled for all row indexes i, k and all columns indexes j, l, which means, that this condition has to be fulfilled by all quaternion of such elements from given matrix, which are positioned in this matrix in rectangle.

In max-plus algebra we can define Monge matrices in the similar way by using operations maximum and plus.

4.4 Usage of the Monge matrices for solving the travelling salesman problem

It is shown that it is possible to solve the travelling salesman problem in polynomial time easily. We assume that setting matrix of this problem is Monge. This problem is studied in [3], [2], [8]. Solving for Monge matrices are based on the pyramidal way concept, which is defined as following:

Let assume that way $\phi = \langle 1, i_1, i_2, \dots, i_r, n, j_1, .., j_{n-r-2} \rangle$ is pyramidal, if following condition holds true:

$$i_1 < i_2 < \cdots < i_r, j_1 > j_2 > \cdots > j_{n-r-2}$$

Then we can formulate following sentence which reduce the computation complexity for this special type of matrices into the quadratic. The problem is studied in [3], where is described the following Theorem.

Theorem 1. The optimal route which is pyramidal exists if the matrix consisting of the costs of the ways between the cities, in max-plus algebra, is Monge. This route can be found by using the dynamic programing with the algorithm of the computation complexity $O(n^2)$.

There can be also other special cases. If distances between all cities have value 1, than computation complexity for solving travelling salesman problem is linear.

4.5 The travelling salesman problem in max-min algebra

The travelling salesman problem in the max-min algebra is the defined analogically as in the max-plus algebra, although the interpretation is different. Graph can be interpreted as a net. The value on each edge represents the size of the minimal flow on this edge. Finding ring connection with maximal throughput is the problem. In max-min algebra analogical statement as Theorem 1**,** can be formulated:

Theorem 2. The optimal route exists if the matrix consisting of the costs of the ways between the cities in max-min algebra is Monge. This route can be found by using the dynamic programing with the algorithm of the computation complexity of $O(n^2)$.

The traveling salesman problem can be solved most effectively for strict Monge matrices. For this special type of matrices the traveling salesman problem can be resolved in the linear time with the algorithm of the computation complexity of O(n). Detailed information, including the Theorem 2, can be found at [3], [8], [10].

Acknowledgements

The research described in the paper was financially supported by project INDOP CZ.1.07/2.2.00/28.0327

References

[1] Apllegate, D. L., Bixby, R. E., Chvátal, V., Cook, W. J. *The Traveling Salesman Problem – A Computational Study*. Princeton University Press, 2006.
[2] Burkard, R., E., Klinz, B., Rudolf, R. Perspectives of monge properties in optimization. In *DiSC.APL.MATH 70,* 1996, pages 95-161.
[3] Burkard, R., Sandholzer, W. Efficiently solvable special casem of bottleneck traveling salesman problems. In *DiSC.APL.MATH 32,* 1991, pages 61-76.
[4] Cuningham-Green, R. A. *Minimax algebra, Lecture notes in Economics and Mathemtaics systems*, 166, Springer-Verlag, Berlin, 1979.
[5] Hynek, J. *Genetické algoritmy a genetické programování.* Grada, Praha, 2008.
[6] Kučera, P. *Metodologie řešení okružního dopravního problému*. ČZU, Praha, 2009.
[7] Lawler, E. L., Lenstra, J. K., Rinnoy Kan, K. H. G., Shmoys, D. B. *The Traveling Salesman Problem.* John Wiley and Sons Ltd., England, 1985.
[8] Park, J. K. A special case of the n-vertex travelling salesman problem that can be sloved in O(n) time. In *INFORM.PROCESS.LETT 40,* 1991, pages 247-254.
[9] Van der Veen, J. A. A. An O(n) algorithm to solve the bottleneck travelling salesman problem stricted to ordered product matrices. In *DiSC.APL.MATH 47*, 1993, pages 57-75
[10] Zelinka, I. *Evoluční výpočetní techniky: principy a aplikace*. BEN, Praha, 2009.

Elimination of Regional Economic Disparities as Optimal Control Problem

Pavel Pražák[1]

Abstract. The paper deals with an issue related to the existence of regions and their economic development. Particularly the problem of possible reduction of the existing economic regional disparities is solved. Regional disparities are understood here only as the difference among the economic productivity of the regions. It is assumed that a part of the public income of the more productive regions is shifted into the area of infrastructure of the less productive regions in order to reduce disparities in regional productivity and differences in incomes between populations of regions. The model that is formulated as optimal control problems in continuous time provides necessary conditions for success of this aim of regional politics and it shows how to reduce the difference between productivity of regions. To find optimal conditions Pontryagin maximum principle is used.

Keywords: region, regional development, regional disparities, redistribution, allocation of investment, optimal control, Pontryagin maximum principle.

JEL classification: R11, C61
AMS classification: 49J15

1 Introduction

The aim of regional policy is regional development aimed at increasing of their coherence and their competitiveness. Because it is more efficient to allocate investment and concentrate economic activity in those regions that have higher productivity, see e.g. [8], regions that have lower productivity may be lagged behind the more efficient regions. Such regions can then become less competitive. For the sake of maintaining cohesion of regions it would be useful to make some redistribution of income between more productive and less productive regions. In order to diminish possible differences between regions and their productivity, governments aimed its investment to infrastructure to the less developed regions. The aim of such efforts is to strengthen the capital stock of the population of the regions, helping to improve the technological level of the region and thus enhance the competitiveness of such regions, cf. [5]. The paper will present a mathematical model that deals with income redistribution and regions investing in their infrastructure. The model is based on ideas in [3], [6] and [7].

2 Problem Description and Model

Consider an economic unit with one central government, which consists of two regions. One of the region is considered to be more developed and its residents will have higher incomes, while the second region is considered to be less developed and its residents will have lower incomes. Quantities describing the richer region will be denoted by index r (rich) and variables related to the poorer region will be denoted by index p (poor). The model is based on the assumptions of free capital and immobile labor, it means that no dramatic migration of workers to work in a richer region is considered.

[1]University of Hradec Kralove, Faculty of Informatics and Management, Department of Informatics and Quantitative Methods, Rokitanského 62, Hradec Králové, e-mail: pavel.prazak@uhk.cz

2.1 Households

Assume that all households in both regions have the same logarithmic utility function and the same discount rate ρ, $\rho \in (0,1)$. This means that the utility functional of representative households in the region i, $i \in \{r,p\}$, is defined as

$$U_i = \int_0^\infty e^{-\rho t} \cdot \ln c_i(t) dt, \tag{1}$$

where $c_i(t)$ represents household consumption in time t, $t \in [0,\infty)$. Assume further that household income in time t is determined by the wage $w_i(t)$ and by inscribed interest from the assets $a_i(t)$ of the household, it means that for the household budget constraint can be written the following equation

$$\dot{a}_i(t) = r_i(t)a_i(t) + w_i(t) - c_i(t), \tag{2}$$

where $r_i(t)$, $r_i(t) \in (0,1)$, is the interest rate in the region i, $i \in \{r,p\}$, at time t, $t \in [0,\infty)$. Households seek to maximize the utility functional (1) subject to the condition (2). This problem is an optimal control problem with state variable $a(\cdot)$ and control variable $c(\cdot)$ that can be solved by Pontryagin maximum principle, see [4]. The Hamiltonian for this problem is

$$H(a_i, c_i, p_i) = e^{-\rho t} \ln c_i + p_i(r_i + w_i - c_i),$$

where we omit variable t denoting time and p_i, $p_i = p_i(t)$, is the adjoint function of the optimal control problem. Observing that H is a concave function in the variable c_i and that it is differentiable we can apply the first-order condition to get

$$\frac{\partial H}{\partial c_i} = e^{-\rho t} \cdot \frac{1}{c_i} - p_i = 0.$$

The adjoint equation is

$$\dot{p}_i = -\frac{\partial H}{\partial a_i} = -p_i a_i.$$

If the result from the first relation is substituted to the adjoint equation it yields

$$\frac{\dot{c}_i(t)}{c_i(t)} = r_i(t) - \rho, \tag{3}$$

where $i \in \{r,p\}$. For the given problem the following transverzality condition can be considered

$$\lim_{t\to\infty} \frac{a_i(t)}{c_i(t)} e^{-\rho t} = 0. \tag{4}$$

Linear ordinary differential equation (2) can be solved by integrating factor method. Employing this method and relations (3) and (4) we finally gain

$$c_i(t) = \rho \cdot (a_i(t) - h_i(t)), \tag{5}$$

where $i \in \{r,p\}$ and $h_i(t)$ is the household wealth defined as

$$h_i(t) = \int_t^\infty w_i(s) \exp\left(-\int_t^s r_i(v)dv\right) ds. \tag{6}$$

2.2 Firms

Consider that in each region i, $i \in \{r,p\}$, there are a large number of competing firms that use private capital K_i and labor L_i for their production Y_i. The aggregate product is then used for consumption and private and public investment. We assume that firms have the same Cobb-Douglas production function

$$Y_i = A_i K_i^\alpha L_i^{1-\alpha}, \tag{7}$$

where A_i represents a general exogenous level of technology in region i that can be partly determined by productive government spending, see [3]. Exponents α and $1-\alpha$ respectively represents the coefficient of elasticity due to the production of capital and due to the production of labor respectively.

In relation to the considerations undertaken it can be assumed that the technological level of production depends on the infrastructure of the region. This dependence, which significantly affects the form of the model, will be specified later. For the time evolution of private capital the standard relationship is assumed

$$\dot{K}_i(t) = I_i(t) - \delta K_i(t), \tag{8}$$

where I_i represents gross investment in private capital in region i, $i \in \{r, p\}$, at the instant t, $t \in [0, \infty)$, and δ represents the rate of capital depreciation. If we consider installation costs of investment and normalize the price per unit of capital the net flow of present value of the company's profit in each region i, $i \in \{r, p\}$, can be expressed as the following functional

$$V_i = \int_0^\infty [(1-\tau)\{A_i(t)K_i(t)^\alpha L_i(t)^{1-\alpha} - w_i(t)L_i(t)\} - I_i(t)] \cdot \exp\left(-\int_0^t r_i(v)dv\right) dt, \tag{9}$$

where τ, $\tau \in (0, 1)$, is the tax rate defined by the central government, $w_i(t)$ is the labor cost per unit of labor in region i at time t and $r_i(t)$ is the interest rate in region i at time t. Firm's aim is to maximize (9) subject to the constraint (8). This problem is an optimal control problem in which the control variables are labor L_i and investment I_i. Capital K_i is the state variable. Now we apply Potryagin maximum principle, see [4]. The Hamiltonian of the given optimal control problem is

$$H(K_i, L_i, I_i, p_i) = [(1-\tau)\{A_i K_i^\alpha L_i^{1-\alpha} - w_i L_i\} - I_i] \cdot \exp\left(-\int_0^t r_i(v)dv\right) + p_i \cdot (I_i - \delta K_i),$$

where we omit variable t denoting time and p_i, $p_i = p_i(t)$, is the adjoint function of the optimal control problem. The maximum principle allows to formulate necessary conditions of the first order for interior optima as follows

$$\frac{\partial H}{\partial I_i} = -\exp\left(-\int_0^t r_i(v)dv\right) + p_i = 0, \tag{10}$$

$$\frac{\partial H}{\partial L_i} = (1-\tau)\{(1-\alpha)A_i K_i^\alpha L_i^{-\alpha} - w_i\} \cdot \exp\left(-\int_0^t r_i(v)dv\right) = 0. \tag{11}$$

The adjoint equation is

$$\dot{p}_i = -\frac{\partial H}{\partial K_i} = -(1-\tau)\alpha A_i K_i^{\alpha-1} L_i^{1-\alpha} \cdot \exp\left(-\int_0^t r_i(v)dv\right) + p_i\delta. \tag{12}$$

Immediately from (11) we gain

$$(1-\alpha)A_i\left(\frac{K_i}{L_i}\right)^\alpha = w_i. \tag{13}$$

If we combine the relations (10) and (12) we gain

$$(1-\tau)\alpha A_i\left(\frac{K_i}{L_i}\right)^{\alpha-1} = r_i + \delta. \tag{14}$$

To interpret these relations we use definition (7) of production functions in both regions. Since

$$\frac{\partial Y_i}{\partial L_i} = A_i(1-\alpha)K_i^\alpha L_i^{-\alpha} = (1-\alpha)A_i\left(\frac{K_i}{L_i}\right)^\alpha = (1-\alpha)\frac{Y_i}{L_i}$$

the condition (13) implies that the unit price of labor is equal to the marginal production of labor. Similarly, since

$$\frac{\partial Y_i}{\partial K_i} = A_i\alpha K_i^{\alpha-1}L_i^{1-\alpha} = \alpha A_i\left(\frac{K_i}{L_i}\right)^{\alpha-1} = \alpha\frac{Y_i}{K_i},$$

the condition (14) implies that the sum of the current interest rate and depreciation of capital is directly proportional to the marginal production of capital. Because between the two considering regions that constitute one economic unit with one central government is a free movement of private capital, the interest rate is the same in both regions, it means that $r_p(t) = r_r(t)$, $t \in [0, \infty)$. The relation (14) then implies that the relation

$$A_r\left(\frac{K_r(t)}{L_r(t)}\right)^{\alpha-1} = A_p\left(\frac{K_p(t)}{L_p(t)}\right)^{\alpha-1} \tag{15}$$

is valid, where $t \in [0, \infty)$.

2.3 Government Sector

The government makes effort to redistribute its income in the amount

$$\tau(Y_r(t) + Y_p(t))$$

to increase the competitiveness of the poorer region. Based on this decision more investment in infrastructure will occur in this region than in the richer region. Public investment in non-production area in this model will not be considered. Let us denote G_i the capital of the region i, $i \in \{r, p\}$ in the form of infrastructure and further let us denote u, $u \in [0, 1]$, the rate of government redistribution. It means that the government will move the extra production of the richer region of the amount $u\tau Y_r(t)$ to a poorer region. If we assume that infrastructure is an immobile capital, we can write a similar differential equation for it as in the case of time evolution of the private capital as follows

$$\dot{G}_r(t) = \tau(1-u)Y_r(t) - \delta G_r(t), \tag{16}$$

and

$$\dot{G}_p(t) = \tau(Y_p(t) + uY_r(t)) - \delta G_p(t), \tag{17}$$

respectively, where δ, $\delta \in (0, 1)$, represents the capital depreciation as has been mentioned earlier.

The last assumption of the model will be a description of how technological progress occurs in the production function (7). In [1] there is the argument that public spending can impact the production in such a way that the production function exhibits constant returns to scale in both the private and the public capital. These considerations are further modified in monograph [2] where it is considered that public expenditures contribute to production in the same extent as labor. In the present model we assume that public expenditures contribute to building infrastructure, that is regarded as public capital, which contributes to the production of the given region. We therefore assume that the infrastructure enriches labor and that it contributes to the production in the same extent as labor, it means that we consider a model of endogenous technological progress enriching labor in the form

$$A_i(t) = AG_i(t)^{1-\alpha}, \tag{18}$$

where A, $A > 0$, is a constant that represents the general level of technology common in both regions. This assumption implies that increasing levels of public spending on infrastructure in the region leads to improve the technological level of the region. This is an essential element of the model. We avoid here to consider a more complex relationship to the technological level. Provided (18) the production function (7) can be written in the form

$$Y_i = AK_i^{\alpha}(G_iL_i)^{1-\alpha}. \tag{19}$$

Now the relation (15) can be rewritten and instead of it we gain the following relation

$$\frac{K_r(t)}{G_r(t)L_r(t)} = \frac{K_p(t)}{G_p(t)L_p(t)}. \tag{20}$$

where $t \in [0, \infty)$. This relation will be used in the following analysis.

3 Relative Level of Backwardness of Less Productive Region

Under the above given assumptions, the productivity of regions depends on the level of regional infrastructure. This infrastructure is immobile and it progress and adapts slowly in time. To see how the value of productivity in a region changes during time, we introduce a relative measure of productivity $\Omega(t)$ of the poorer region with respect to richer region as the ratio

$$\Omega(t) = \frac{y_p(t)}{y_r(t)} = \frac{Y_p(t)}{L_p(t)} \cdot \left(\frac{Y_r(t)}{L_r(t)}\right)^{-1}. \tag{21}$$

To see how this variable evolve in time and how it is affected by a government redistribution of its income we further introduce relative growth rate of productivity $\gamma_\Omega(t)$ of the poorer region in relation to the richer region

$$\gamma_\Omega(t) = \frac{\dot{\Omega}(t)}{\Omega(t)}. \tag{22}$$

To be able to work with this concept we need to connect it with relations (20) that represents optimal condition for behavior of firms in the economy and (16), (18) respectively, that represents government redistribution of its income. To simplify calculations we introduce the following ratios

$$\kappa(t) = \frac{K_p(t)}{K_r(t)}, \lambda(t) = \frac{L_p(t)}{L_r(t)}, \gamma(t) = \frac{G_p(t)}{G_r(t)}, \tag{23}$$

where $t \in [0, \infty)$. Now (20) can be written as

$$\kappa(t) = \lambda(t) \cdot \gamma(t) \tag{24}$$

and instead of (21) it is possible to write

$$\Omega(t) = \kappa(t)^{\alpha} \cdot \lambda(t)^{-\alpha} \cdot \gamma(t)^{1-\alpha}. \tag{25}$$

To simplify notation we again omit the variable t in further considerations. From (25) we immediately gain

$$\gamma_{\Omega} = \frac{\dot{\Omega}}{\Omega} = \alpha\frac{\dot{\kappa}}{\kappa} - \alpha\frac{\dot{\lambda}}{\lambda} + (1-\alpha)\frac{\dot{\gamma}}{\gamma}. \tag{26}$$

To simplify this relation we use (24) that allows us to find that

$$\frac{\dot{\kappa}}{\kappa} = \frac{\dot{\lambda}}{\lambda} + \frac{\dot{\gamma}}{\gamma}. \tag{27}$$

If we substitute this result (27) into (26) and if we use the meaning of the variable γ given in (23) we find

$$\gamma_{\Omega} = \frac{\dot{\gamma}}{\gamma} = \frac{\dot{G}_p}{G_p} - \frac{\dot{G}_r}{G_r}. \tag{28}$$

Now we can apply relations (16) and (18). Substituting them into () we gain the relation

$$\gamma_{\Omega} = \tau\left(\frac{Y_p + uY_r}{G_p}\right) - \tau(1-u)\frac{Y_r}{G_r}. \tag{29}$$

Using (19) for the poor and the rich region we can rewrite this relation in the following way

$$\gamma_{\Omega} = \tau A\left(\left(\frac{K_p}{G_pL_p}\right)^{\alpha} L_p + u\left(\frac{K_r}{G_rL_r}\right)^{\alpha}\frac{G_rL_r}{G_p} - (1-u)\left(\frac{K_r}{G_rL_r}\right)^{\alpha} L_r\right). \tag{30}$$

If we finally use (19) again we find the resulting relation

$$\gamma_{\Omega} = \tau A\left(\frac{K_p}{G_pL_p}\right)^{\alpha}\left(L_p - L_r + uL_r\left(\frac{G_r}{G_p} + 1\right)\right). \tag{31}$$

3.1 Interpretation

Let us briefly recall that relation (31) was derived for an economic unit consisting of two regions in which the same interest rate exists and the same general level of technology is applied. It means that both regions may borrow funds under the same conditions and have access to the same technologies. Furthermore, the model assumes that the population of poorer region don't massively move to work in the rich region. The following considerations concern only the relationship of productivity in the poorer and richer region and they do not relate to the aggregate growth of output in the given economic unit.

- Let $u = 0$, it means that there is not a redistribution of the total income of the economic unit. The productivity of less developed region can increase if $L_p > L_r$, cf. relation (31). It means that the poorer region can use more labor than the richer region. If $L_p \approx L_r$ then private capital and its using can't contribute to the increasing of productivity of the less developed region.
- Relation (31) implies that $\gamma_{\Omega} > 0$ if and only if

$$u > \frac{L_r - L_p}{L_r} \cdot \frac{G_p}{G_r + G_p}. \tag{32}$$

  Since the expression on the right hand side of the above relation is less than one there is always such u, $u \in [0, 1]$, that (32) is valid. It means that the government can always find a rate of redistribution u of its aggregate income that can increase the productivity of less developed region.

4 Conclusion

In the management of economic units that are divided into regions the central government is facing different problems and challenges. On the one hand it pursues economic goals when it invests into the most productive regions, because such investments are most effective. In this way, however, disparities between different regions may rise and less productive regions may lose their competitiveness. On the other hand, the government pursues political goals, which attempts to eliminate differences in economic development of regions to prevent dissatisfactions of the population in less developed regions. Such dissatisfactions could lead to a reduction of cohesion among regions. In order to eliminate this phenomena the central government can increase the competitiveness of less developed regions when it invests part of their income from the more developed regions into the infrastructure of less developed regions. In the present paper an analytical models that seeks to describe such behavior of central government and to show the effectiveness of such decisions were presented. The model was formulated as an optimal control problem and Pontryagin maximum principle was used to find the necessary conditions of optimal solution. It has been shown that under the assumptions of the model the redistribution of government income can increase the productivity of less productive and less developed region. We can conclude that the politics of redistribution is under the given assumptions efficient.

Acknowledgements

The support of the Czech Science Foundation #402/09/0405 is kindly acknowledged.

References

[1] Barro, J., Government spending in a simple model of endogeneous growth. *Journal of Political Economy 98,* **5,** (Oct., 1990), 103–125.

[2] Barro, R. J., Salla-i-Martin, X., *Economic Growth,* 2nd ed.: MIT Press, London, 2004.

[3] Funke, M., Strulik, H., Growth and convergence in a two-region model: The hypothetical case of Korean unification. *Journal of Asian Economics 16,* **2,** (Apr., 2005), 255–279.

[4] Joffe A. D., Tichomirov V. M. *Teorija ekstremalnych zadač*, Nauka, Moscow, 1974. (in Russian)

[5] 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.

[6] Ono, Y., Shibata, A., Spill-over effects of supply-side changes in two-country economy with capital accumulation. *Journal of International Economics 33,* **1/2,** (Aug., 1992), 127–146.

[7] Pražák P., Matematický model regionálního přerozdělování. In: *Mezinárodní vědecká konference Hradecké ekonomické dny 2011, 1. díl,* Gaudeamus, Hradec Králové, 2011, 270–275. (in Czech)

[8] Takayama A., *Regional allocation of investment: A Further Analysis,* The Quarterly Journal of Economics, **81,** No. 2 (May, 1967), 330 – 337.

Forecasting Financial Time Series

Peter Princ[1], Sára Bisová[2], Adam Borovička[3]

Abstract. Density forecast is an estimate of the probability distribution of the possible future values of a random variable. In our research we compare two different approaches for one-step ahead forecasting in stock market indices with parameters from normal and two-piece normal (TPN) distribution. We analyze three stock market indices S&P 500 from New York Stock Exchange, FTSE 100 from London Stock Exchange and PX from Prague Stock Exchange with daily returns data. For improving density forecasts, the parameters of two-piece normal distribution for approximating the asymmetry (negative skewness) in standardized residuals were estimated by the maximum likelihood method. Rolling one-step ahead predictions for the last 200 observations of the future returns using samples from stock market indices time series were made using the appropriate ARMA-GARCH models with two different density forecasting distributions (normal and TPN). Using the parameters from two-piece normal, estimated on rolling samples, the ARMA-GARCH-TPN predictions were obtained. We compare these two approaches of forecasting by using chosen tests.

Keywords: financial time series, asymmetry, volatility, two-piece normal distribution

JEL Classification: C58
AMS Classification: 91B84

1 Introduction

Volatility indicates a period of time series connected with high variability or growing variance [1, 2, 3, 7, 8]. This phenomenon plays a great role in the process of modelling and analysing of financial time series. It is the main element of the procedure of quantifying the general risk of financial assets. Modelling and forecasting volatility are generally used in investment decision process for capturing suitable risk of potential investment portfolio and in the analysis of VaR model or option price derivation [3].

A large number of literature in econometrics has focused on evaluating the forecast accuracy of volatility models [4,5]. In our study we compare two different approaches for modelling financial time series and its volatility using stock market indices - S&P 500 from New York Stock Exchange, FTSE 100 from London Stock Exchange and PX from Prague Stock Exchange. During the process asymmetric features in these daily financial times series must be faced. In terms of our approach, asymmetries in the conditional distributions by employing block bootstrap procedure with conditional distribution were captured and a comparison between forecasting from constant and rolling sample were involved as well. These two approaches via one-step-ahead forecasts of future returns using samples from mentioned stock market indices time series were compared.

The paper is structured as follows. In Section 2, we briefly discuss the volatility models. Section 3 introduces the two-piece normal distribution. In Section 4, we apply maximum likelihood method for estimating parameters of GARCH models with symmetric and asymmetric distribution on stock market indices mentioned above. Section 5 concludes.

2 Volatility Models

Volatility measured by the standard deviation or variance of returns often represents the risk measurement concept in the financial market. There are many given models for modelling and forecasting volatility in the financial time series. As the volatility is usually conditional, the presence of conditional heteroscedasticity in the econometrics models may be inferred. In other words, the residuals of linear regression model embody variable variance and, in addition, the variance of residuals is affected by their past values.

[1] University of Economics, Prague, Dept. Of Econometrics, W. Churchill Sq. 4, Prague 3, peter.princ@vse.cz
[2] University of Economics, Prague, Dept. Of Econometrics, W. Churchill Sq. 4, Prague 3, xbiss00@vse.cz
[3] University of Economics, Prague, Dept. Of Econometrics, W. Churchill Sq. 4, Prague 3, adam.borovicka@vse.cz

Firstly we define the basic volatility model developed by R. F. Engle, ARCH(p). Model can be written as follows [9]

$$r_t = \Phi_0 + \Phi_1 r_{t-1} + \Phi_2 r_{t-2} + ... + \Phi_s r_{t-s} + \varepsilon_t \tag{1a}$$

$$\sigma_t^2 = \omega + a_1 \varepsilon_{t-1}^2 + a_2 \varepsilon_{t-2}^2 + ... + a_p \varepsilon_{t-p}^2 \tag{1b}$$

$$\varepsilon_t = \sigma_t e_t, \tag{1c}$$

where $\{\varepsilon_t\}$ is the conditional heteroskedastic process, or it is the variance of the shocks time variable and depends on the past *p* shocks $\varepsilon_{t-1},...,\varepsilon_{t-p}$, the conditional variance σ_t^2 is set on the basis of the information accessible in time *t*-1, e_t is distributed with standard normal distribution *N*(0,1), $\Phi_0, \Phi_1, ..., \Phi_s, \alpha_1, ..., \alpha_p, \omega$ are the parameters of the model, where $\alpha_1,...,\alpha_p \geq 0$, $\omega > 0$ conditions ensure the positivity of the conditional variance.

In many cases, ARCH model requires an estimation of many parameters *p*, hence GARCH model was developed by T. P. Bollerslev. As we can see in [2], the improvement of the process lies in adding of lagged conditional variance, and therefore the GARCH model accepts the dependency of conditional variance on previous own lags. Instead of (1b), we can formulate for GARCH (p,q)

$$\sigma_t^2 = \omega + \sum_{p=1}^{P} \alpha_p \varepsilon_{t-p}^2 + \sum_{q=1}^{Q} \beta_p \sigma_{t-q}^2, \tag{2}$$

where *Q* lags of the conditional variance $\sigma_{t-1}^2, \sigma_{t-2}^2, ..., \sigma_{t-Q}^2$ are included. In both models, the kurtosis of shocks ε_t is greater than the kurtosis of normal as is proved in [9].

Similar to [2], we employ *maximum likelihood method* in order to estimate the volatility model.

3 Two-piece normal distribution

According to [10], if a random variable *X* has a two piece normal distribution, $X \sim SN(\mu, \lambda^2, \tau^2)$, then its probability density function is

$$f(x) = \begin{cases} c \exp\left[-\frac{1}{2\lambda^2}(x-\mu)^2 \right] & \text{if } x \leq \mu \\ c \exp\left[-\frac{1}{2\tau^2\lambda^2}(x-\mu)^2 \right] & \text{if } x > \mu, \end{cases} \tag{3}$$

where $c = \sqrt{2/\pi}\lambda^{-1}(1+\tau)^{-1}$.

The density of the $SN(\mu, \lambda^2, \tau^2)$ distribution is proportional to the density of the $N(\mu, \lambda^2)$ distribution to the left of the mode μ, to the right of the mode it is proportional to the density of the $N(\mu, \tau^2\lambda^2)$ distribution. This described probability distribution is skewed to the left for $\tau < 1$, to the right for $\tau > 1$ and for $\tau = 1$ it reduces to the usual normal distribution.

4 Application

Three stock market indices S&P 500 from New York Stock Exchange, FTSE 100 from London Stock Exchange and PX from Prague Stock Exchange were chosen for this study. All times series from 3 January 1995 to 27 April 2012 were considered. In order to eliminate the non-stationarity in selected financial times series, we compute the differences between the logarithms of the prices on adjacent days

$$w_t = \ln z_t - \ln z_{t-1} = \ln \frac{z_t}{z_{t-1}}, \tag{4}$$

where z_t represents the closing price of the particular daily index in period *t*.

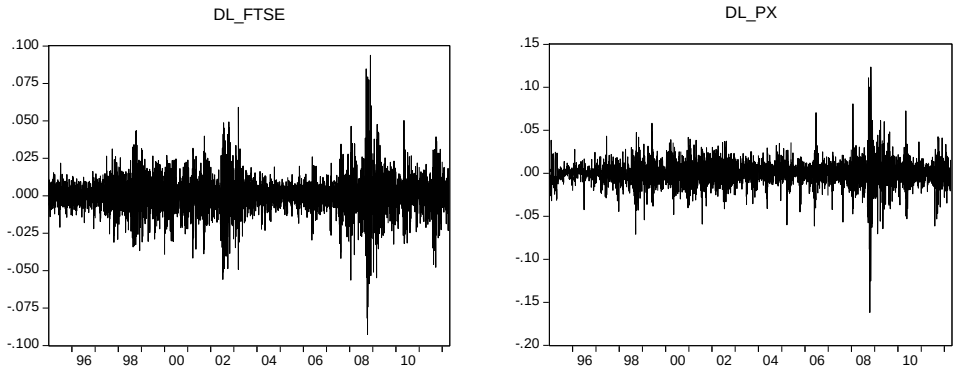


Figure 1 Daily FTSE 100 and PX returns from January 1995 to April 2012

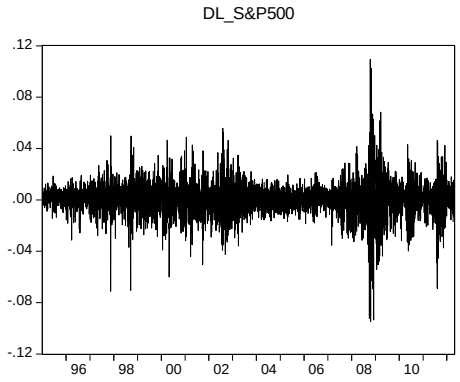


Figure 2 Daily S&P 500 returns from January 1995 to April 2012

Statistic	FTSE 100 returns	PX returns	S&P 500 returns
Observations	4374	4374	4374
Mean	0.00014486	0.000081335	0.00025752
Std. Deviation	0.012249	0.014421	0.012757
Skewness	-0.13847	-0.42767	-0.23466
Excess Kurtosis	5.6562	11.252	7.6101
Minimum	-0.092646	-0.16185	-0.094695
Maximum	0.093842	0.12364	0.10957

Table 1 Descriptive statistics for daily FTSE 100, PX and S&P 500 returns

In the previous graphs (Figure 1 and 2) and table (Table 1), daily returns of each index are slightly negatively skewed and positively biased.

Hua [6] studied two piece normal distribution for obtaining more accurate predictions for symmetric volatility model ARMA-GARCH, where estimated parameters of two-piece normal distribution (TPN) were obtained from residuals received from the mean equation of ARMA-GARCH process. The maximum likelihood method and stationary bootstrap is employed in his study for estimation of TPN. One-step ahead forecast of conditional mean and standard deviations of ARMA-GARCH process with TPN is defined by the formulas below.

Conditional mean of one-step ahead forecast

$$\hat{y}_{t+1} = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\hat{\mu}_{\eta}. \tag{5}$$

Conditional standard deviations of one-step ahead forecast

$$\hat{\sigma}_{t+1,1} = \hat{\sigma}_{t+1}\hat{\sigma}_{t+1,\eta 1}, \qquad \hat{\sigma}_{t+1,2} = \hat{\sigma}_{t+1}\hat{\sigma}_{t+1,\eta 2}. \tag{6}$$

Where $\hat{\mu}_{t+1}$ and $\hat{\sigma}_{t+1}$ are the conditional mean and standard deviation of the one-step ahead forecast of ARMA-GARCH model and $\hat{\mu}_{\eta}$ is the mode of two-piece normal distribution, $\hat{\sigma}_{t+1,\eta 1}$ is the standard deviation of the one-step ahead forecast if $x \le \hat{\mu}_{\eta}$ and $\hat{\sigma}_{t+1,\eta 2}$ is the standard deviation of the one-step ahead forecast if $x > \hat{\mu}_{\eta}$. Considering the formula (3) we transform parameters of TPN distribution with this approach $\hat{\sigma}_{t+1,\eta 1} = \lambda$ and $\hat{\sigma}_{t+1,\eta 2} = \tau\lambda$.

In his study [6] Hua concludes that this estimation method provides more accurate estimates for one-step ahead forecast in the financial time series. This proposition is based on the fact that in the case of skewed time series the mode is a better measure of central tendency the mean. In the next section we examine these conclusions.

The returns of indices are stationary under the condition of ADF unit root test [1]. We scrutinise conditional heteroscedasticity by ARCH test and conditional heteroscedasticity of nonlinear type by common SB, PSB, NSP tests [1]. Results from these tests are in Table 2 and Table 3.

Index	t-statistic	Test critical values at 1% level	Prob.*
FTSE 100	-30.31201	-3.431663	0.0000
PX	-60.22226	-3.431675	0.0001
S&P500	-50.97834	-3.431666	0.0001

*MacKinnon (1996) one-sided p-values.
Notes: Test critical values -3.43 (1% level), -2.86 (5% level), 2.57 (10% level)

Table 2 ADF unit root tests

Index	LM test statistic (TR^2)	Critical value ($\chi^2_{0.95}(3)$)
FTSE 100	3183.103	7.8147
S&P 500	3234.073	7.8147
PX	2747.082	7.8147

Notes: Null hypothesis - residuals (conditional mean equation) do not include asymmetry of any type

Table 3 Common SB, PSB, NSB test of asymmetry in conditional heteroscedasticity

According to the results of heteroscedasticity tests, the conditional heteroscedasticity is present in each time series and depends on the size of positive and negative returns, where level of negative shocks affects the conditional heteroscedasticity slightly significantly than the positive shocks do.

We estimate most suitable ARMA-GARCH model for each stock index, for FTSE 100 AR(3)-GARCH(1,1), for S&P 500 ARMA(1,1)-GARCH(2,2) and for PX AR(1),AR(4)-GARCH(1,1), where the constant term is included in each conditional mean equation. The histograms of standardised residuals with skewness -0.495299 (S&P 500), -0.277981 (FTSE 100) and -0.242310 (PX) are in Figure 3.

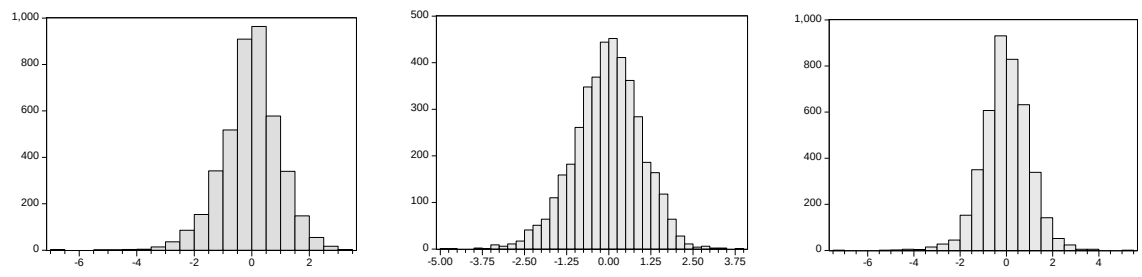


Figure 3 Histograms of standardized residuals for S&P 500, FTSE 100 and PX

Point and interval predictions, based on chosen models, for the last 200 periods (ex post prediction) with rolling sample were constructed. The values of estimated parameters were closed to the parameters gained from the bootstrapped data. All the predictions for ARMA-GARCH-TPN were computed according to formulas (5) - (6).

We employ Q-Q plots and Kolgomorov-Smirnov test for comparison of the quality of ex post prediction between ARMA-GARCH and ARMA-GARCH-TPN. Results from these tests are in the Table 4.

Index: Method	Kolgomorov-Smirnov test statistic	Asymp. *p*-value
FTSE 100: GARCH	0.4545	1.1570e-018
FTSE 100: GARCH_TPN	0.5354	1.0056e-025
S&P 500: GARCH	0.4141	1.4504e-015
S&P 500: GARCH_TPN	0.5354	1.0056e-025
PX: GARCH	0.3939	4.0040e-014
PX: GARCH_TPN	0.4394	1.8151e-017

Notes: Null hypothesis - empirical and fitted values are prom the same continuous distribution, two-tailed alternative hypothesis.

Table 4 Two-sample Kolgomorov-Smirnov test for rolling sample

Q-Q plot describes the relation between quantiles of the empirical values represented by real-time returns and the point predictions of selected model of each time series. In all cases, the ARMA-GARCH predictions are slightly closer to empirical values. These plots are displayed in the Figure 4 and Figure 5.

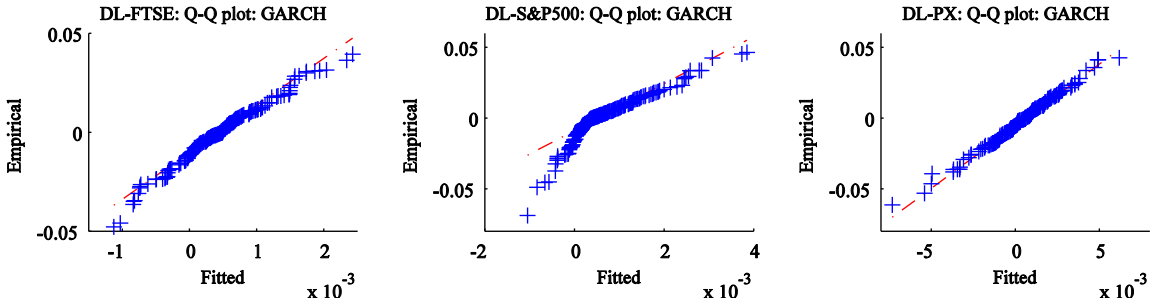


Figure 4 Q-Q plots for one-step ahead forecasts of GARCH with normal distribution for FTSE 100, S&P 500 and PX

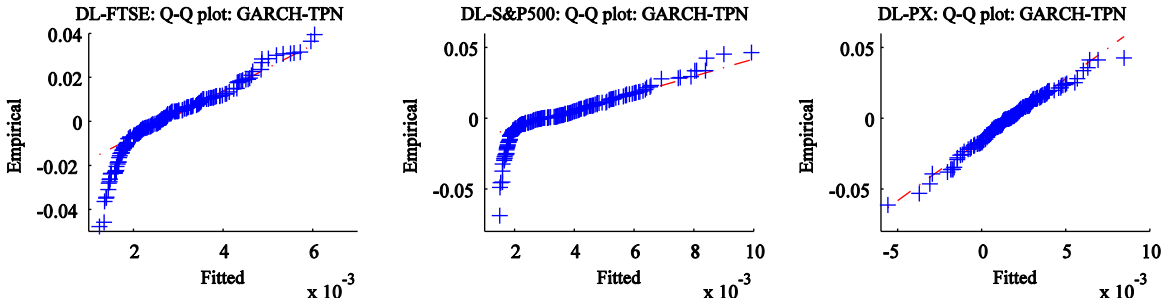


Figure 5 Q-Q plots for one-step ahead forecasts of GARCH with two-piece normal distribution for FTSE 100, S&P 500 and PX

Similar to Hua [6], we applied stationary bootstrap to the standardised residuals from the conditional mean equation and bootstrapped 4 millions data from which the parameters of TPN were estimated by the maximum likelihood estimation and then one-step ahead forecasts for the last 200 periods were conducted. We compared the two approaches of predicting - one-step ahead forecasting from the bootstrapped data sample and appropriate forecasts from the rolling sample. The point and interval predictions were transformed according to formulas (5) - (6). The results from both techniques are only slightly different. Table 5 reports the estimated parameters of TPN distribution obtained from the bootstrapped data.

Index	Mode	Sigma_1	Tau	Sigma_2
FTSE 100	0.169464	1.12507	0.768082	0.86414602
	(0.001187)	(0.000722)	(0.001013)	
S&P 500	0.193028	1.147181	0.72654	0.83347288
	(0.001133)	(0.000582)	(0.000874)	
PX	0.075277	1.066143	0.869876	0.92741221
	(0.000989)	(0.000488)	(0.000800)	

Table 5 Estimated coefficient by MLE for two-piece normal distribution

5 Conclusion

This study deals with modelling and forecasting of three stock indices. Firstly, the condition mean equations were constructed using the linear stationary ARMA processes of Box-Jenkins methodology. Consequently, the appropriate models of symmetric conditional variance (GARCH models) were estimated. For improved density forecasts, the parameters of TPN distribution for approximating the asymmetry (negative skewness) in standardised residuals were estimated by the maximum likelihood method. Rolling one-step-ahead predictions for the last 200 observations were made using the appropriate ARMA-GARCH models. Using the TPN parameters, estimated on rolling samples, the ARMA-GARCH-TPN predictions were obtained. We scrutinise the method developed by Hua [6] and conclusions based on our results are on the contrary to Hua.

Especially in the case of FTSE 100 index, using the TPN transformation, the received predictions fit the empirical values worse than that obtained by using standard ARMA-GARCH models. The main idea of mentioned approach is that the mode should be more relevant and representative measure of central tendency than the mean when studying asymmetric distributions. In our opinion, this methodology could be more convenient in situations, where the skeweness is caused by particular outliers, which is not the case of our samples. The observations are cumulated around the zero value, with several slightly outlying values on both sides (especially during the year 2008 reflecting the financial crisis). Moreover, in finance, the investors usually prefer slightly underestimated yields than underestimated possible losses. However, this method fits more values close to mode, the negative values move further away which could brings huge losses in the case of unexpected negative shocks and it is a drawback in financial analysis.

Acknowledgements

The research project was supported by Grant No. IGA F4/16 /2011, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Arlt, J., Artlová, M.: *Ekonomické časové řady.* Professional Publishing, Praha, 2009.
[2] Brooks, Ch.: *Introductory Econometrics in Finance*. Cambridge University Press, New York, 2008.
[3] Cipra, T.: *Finanční ekonometrie.* EKOPRESS, Praha, 2008.
[4] Diebold, F. X., Mariano R. S.: Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13** (1995), 253-263.
[5] Hansen, P. R., Lunde, A. (2003). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)?. *Journal of Applied Econometrics*, **20** (2005), 873-889.
[6] Hua, Z. S., Zhang, B.: Improving density forecast by modelling asymmetric features: An application to SP&500 returns. *European Journal of Operational Research* **185** (2008), 716-725.
[7] Hušek, R.: *Ekonometrická analýza*. Oeconomica, Praha, 2007.
[8] Hušek, R.: *Aplikovaná ekonometrie: teorie a praxe.* Oeconomica, Praha, 2009.
[9] Sheppard, K.: *Financial Econometrics Notes*. University of Oxford, Oxford, 2010.
[10] Villani, M., Larsson, R.: The Multivariate Split Normal Distribution and Asymmetric Principal Components Analysis. *Communications in Statistics – Theory and Methods* **35** (2006), 1123-1140.

Measuring transitivity of fuzzy pairwise comparison matrix

Jaroslav Ramík[1]

Abstract. A pair-wise comparison matrix is the result of pair-wise comparison a powerful method in multi-criteria optimization. When comparing two elements, the decision maker assigns the value representing the element of the pair-wise comparison matrix. In AHP, the matrix represents a multiplicative preference relation. Here, consistency property plays an essential role. To provide a consistency measure of the pair-wise comparison matrix, the consistency ratio is defined in AHP. In some situations another interpretation is convenient. The preferences can be represented by a fuzzy preference relation, given by the membership function denoting the preference degree (or intensity) of one alternative over the other. The role of consistency is played by the concept of transitivity. In this paper we investigate relations between several types of transitivity of fuzzy relations and multiplicative preference relations. Similarly to the consistency ratio we also define the grade of transitivity. Consequently, we obtain corresponding priority vectors. An illustrative numerical example is supplemented.

Keywords: multi-criteria optimization, pair-wise comparison matrix, AHP

JEL classification: C44
AMS classification: 90B50, 90C29, 91B08

1 Introduction

The pair-wise comparison matrix is a powerful inference tool in decision making (DM), see e.g. [3], [6], that can be also used as a knowledge acquisition technique for knowledge-based systems. It is useful for assessing the relative importance of several objects, when this cannot be done by direct rating. As it is known, most decision processes are based on preference relations, in the sense that they are linked to some degree of preference of any alternative over another. Therefore, to establish properties to be verified by such preference relations is very important for designing good DM models. Three of these properties are investigated in this paper, it is so called reciprocity, consistency property and transitivity property. The lack of consistency or transitivity in DM can lead to wrong conclusions, see [6]. That is why it is important to study conditions under which consistency and/or transitivity is satisfied [3]. On the other hand, perfect consistency/transitivity is difficult to obtain in practice, particularly when evaluating preferences on a set with a large number of alternatives. Then it is important to know, whether and in what grade our preferences are coherent each other. In other words, we ask how strongly are our preferences consistent, or transitive. Hence, our goal here is to derive some simple tools enabling us to measure the grade of consistency/transitivity of pair-wise comparison relations, or, giving us some information about inconsistency of our preferences, i.e. how much consistency/transitivity of our preferences is damaged. As a consequence, we obtain also priority vectors that are used for ranking the variants in DM processes.

2 Multiplicative and additive preferences

The problem can be formulated as follows. Let $X = \{x_1, x_2, ..., x_n\}$ be a finite set of alternatives. These alternatives have to be classified from best to worst, using the information given by a decision maker. However, it can often be difficult for the decision maker to express exact estimates of the ratios of importance of alternatives. This leads us to assume that the preferences over the set of alternatives,

[1]Silesian University in Opava, School of Business Administration in Karvina, Department of Math. Methods in Economics, University Sq. 1934/3, 73340 Karvina, Czech Republic, ramik@opf.slu.cz

X, may be represented (at least) in the following two ways: multiplicative and additive. Let us assume that the preferences on X are described by a preference relation on X given by a positive $n \times n$ matrix $A = \{a_{ij}\}$, where $a_{ij} > 0$ for all i, j indicates a preference intensity for alternative x_i to that of x_j, i.e. it is interpreted as "x_i is a_{ij} times as good as x_j". According to [6], T. Saaty suggested measuring a_{ij} using a ratio scale, particularly the scale $\{1/9, 1/8, ..., 1, ..., 8, 9\}$. The elements of $A = \{a_{ij}\}$ satisfy the following reciprocity condition [6].

A positive $n \times n$ matrix $A = \{a_{ij}\}$ is *multiplicative-reciprocal* (*m-reciprocal*), if

$$a_{ij}.a_{ji} = 1 \text{ for all } i, j. \tag{1}$$

A positive $n \times n$ matrix $A = \{a_{ij}\}$ is *multiplicative-consistent* (or, *m-consistent*) [3], [6], if

$$a_{ij} = a_{ik}.a_{kj} \text{ for all } i, j, k. \tag{2}$$

Notice that $a_{ii} = 1$ for all i, and also (2) implies (1), i.e. an m-consistent positive matrix is m-reciprocal (however, not vice-versa). Then, (2) can be rewritten equivalently as

$$a_{ik}.a_{kj}.a_{ji} = 1 \text{ for all } i, j, k. \tag{3}$$

Notice that here $a_{ij} > 0$ and m-consistency is not restricted to the Saatys scale. The above mentioned interpretation of preferences on X described by a positive matrix is, however, not always appropriate for a decision maker. Evaluating the preference of two elements of a pair, say, x_i and x_j with respect to e.g. "design of a product" might cause a problem. Here, saying e.g. that x_i is 3 times as good as x_j is peculiar. Using word categories, e.g. "moderately (strongly, very strongly etc.) better", as it is recommended in AHP [6], is not way out. A more natural way seems to be the following: divide 100% of the property in question into two parts and then assign the first part to the first element and the rest to the second one, see [4]. In other words, when comparing x_i to x_j the decision maker assigns the value b_{ij} to x_i and b_{ji} to x_j, whereas $b_{ij} + b_{ji} = 1$ (i.e. 100%). With this interpretation, the preferences on X can be represented also by a *fuzzy preference relation*, with membership function $\mu_R : X \times X \to [0; 1]$, where $\mu_R(x_i, x_j) = b_{ij}$ denotes the preference of the alternative x_i over x_j [3], [4], [6]. Important properties of the above mentioned matrix $B = \{b_{ij}\}$ can be summarized as follows.

An $n \times n$ matrix $B = \{b_{ij}\}$ with $0 \leq b_{ij} \leq 1$ for all i and j is *additive-reciprocal* (*a-reciprocal*) [1], if

$$b_{ij} + b_{ji} = 1 \text{ for all } i, j. \tag{4}$$

Evidently, if (4) holds for all i and j, then $b_{ii} = 0.5$ for all i.

For making a coherent choice (when assuming additive fuzzy preference matrices) a set of properties to be satisfied by such relations have been suggested in the literature [3], [7].

Transitivity is one of the most important properties concerning preferences, and it represents the idea that the preference intensity obtained by comparing directly two alternatives should be equal to or greater than the preference intensity "between" those two alternatives obtained using an indirect chain of alternatives. Some of the suggested properties – candidates for consistency property – are given here, see also [1], [7].

Let $B = \{b_{ij}\}$ be an $n \times n$ *a*-reciprocal matrix with $0 \leq b_{ij} \leq 1$ for all i and j.

1. *Weak transitivity* [7]:

$$\text{If } b_{ij} \geq 0.5 \text{ and } b_{jk} \geq 0.5 \text{ then } b_{ik} \geq 0.5 \text{ for all } i, j, k. \tag{5}$$

2. *Restricted max-transitivity* [7]:

$$\text{If } b_{ij} \geq 0.5 \text{ and } b_{jk} \geq 0.5 \text{ then } b_{ik} \geq max\{b_{ij}, b_{jk}\} \text{ for all } i, j, k. \tag{6}$$

3. *Multiplicative-transitivity (m-transitivity)* [7]:

$$b_{ij}.b_{jk}.b_{ki} = b_{ik}.b_{kj}.b_{ji} \text{ for all } i, j, k. \tag{7}$$

If $b_{ij} > 0$ for all i and j, then (7) can be rewritten as

$$\frac{b_{ij}}{b_{ji}}.\frac{b_{jk}}{b_{kj}}.\frac{b_{ki}}{b_{ik}} = 1 \text{ for all } i, j, k. \tag{8}$$

It is easy to prove that if $B = \{b_{ij}\}$ is m-transitive, then it is restricted max-transitive. Evidently, the opposite is not true. Notice that if B is m-consistent then B is m-transitive. Moreover, if $B = \{b_{ij}\}$ is m-reciprocal, then B is m-transitive iff B is m-consistent.

4. *Additive-transitivity* (*a-transitivity*) [7]:

$$b_{ij} + b_{jk} + b_{ki} = 1.5 \text{ for all } i, j, k. \tag{9}$$

It is not difficult to prove that if $B = \{b_{ij}\}$ is additive transitive, then it is restricted max-transitive. Therefore, the additive transitivity is stronger than restricted max-transitivity.

3 Additive versus multiplicative-reciprocal matrices

In this section we shall investigate some relationships between a-reciprocal and m-reciprocal pair-wise comparison matrices. We start with extension of the result published by E. Herrera-Viedma et al. [1]. For this purpose, given $\sigma > 1$, we define the following function φ_σ and its inverse function φ_σ^{-1} as

$$\varphi_\sigma(t) = \frac{1}{2}(1 + \frac{\ln t}{\ln \sigma}) \text{ for } t \in [1/\sigma; \sigma], \tag{10}$$

$$\varphi_\sigma^{-1}(t) = \sigma^{2t-1} \text{ for } t \in [0; 1]. \tag{11}$$

We obtain the following results, characterizing a-transitive and m-consistent matrices, see [1].

Proposition 1. Let $A = \{a_{ij}\}$ be an $n \times n$ matrix with $\frac{1}{\sigma} \leq a_{ij} \leq \sigma$ for all i and j.

If $A = \{a_{ij}\}$ is m-consistent then $B = \{\varphi_\sigma(a_{ij})\}$ is a-transitive.

Proposition 2. Let $B = \{b_{ij}\}$ be an $n \times n$ matrix with $0 \leq b_{ij} \leq 1$ for all i and j.

If $B = \{b_{ij}\}$ is a-transitive then $A = \{\varphi_\sigma^{-1}(b_{ij})\}$ is m-consistent.

Now, let us define the function ϕ and its inverse function ϕ^{-1} as follows

$$\phi(t) = \frac{t}{1+t} \text{ for } t > 0, \tag{12}$$

$$\phi^{-1}(t) = \frac{t}{1-t} \text{ for } 0 < t < 1. \tag{13}$$

Proposition 3.

Let $A = \{a_{ij}\}$ be an $n \times n$ matrix with $0 < a_{ij}$ for all i and j.

If $A = \{a_{ij}\}$ is m-consistent then $B = \{b_{ij}\} = \{\phi(a_{ij})\}$ is m-transitive.

Proposition 4.

Let $B = \{b_{ij}\}$ be an a-reciprocal $n \times n$ matrix with $0 < b_{ij} < 1$ for all i and j.

If $B = \{b_{ij}\}$ is m-transitive then $A = \{a_{ij}\} = \{\phi^{-1}(b_{ij})\}$ is m-consistent.

From Proposition 2 it is clear that the concept of m-transitivity plays a similar role for a-reciprocal matrices as the concept of m-consistency does for m-reciprocal matrices. That is why it is reasonable to introduce the following definition: Any $n \times n$ nonnegative a-reciprocal matrix $B = \{b_{ij}\}$ which is m-transitive is called *additively consistent* (*a-consistent*). Then Proposition 4 can be reformulated accordingly.

In practice, perfect consistency/transitivity is difficult to obtain, particularly when measuring preferences on a set with a large number of alternatives.

4 Inconsistency of pair-wise comparison matrices, priority vectors

If for some positive $n \times n$ matrix $A = \{a_{ij}\}$ and for some $i, j, k = 1, 2, ..., n$, multiplicative consistency condition (2) does not hold, then A is said to be *multiplicative-inconsistent* (or, *m-inconsistent*). Eventually, if for some $n \times n$ matrix $B = \{b_{ij}\}$ with $0 \leq b_{ij} \leq 1$ for all i and j, and for some $i, j, k = 1, 2, ..., n$, (7) does not hold, then B is said to be *additive-inconsistent* (or, *a-inconsistent*) . Finally, if for some $n \times n$ matrix $B = \{b_{ij}\}$ with $0 \leq b_{ij} \leq 1$ for all i and j, and for some $i, j, k = 1, 2, ..., n$, (9) does not hold, then B is said to be *additive-intransitive* (or, $a-intransitive$). In order to measure the grade of inconsistency/intransitivity of a given matrix several measurement methods have been proposed in the literature. In AHP, multiplicative reciprocal matrices have been considered, see [6].

As far as additive-reciprocal matrices are concerned, some methods for measuring a-inconsistency/a-intransitivity are proposed in this section. Here, instead of positive matrices we consider preference matrices with nonnegative elements, i.e. some elements are eventually zeros. Measuring inconsitency of such matrix is based on Perron-Frobenius theory which is known in several versions, see [2]. The Perron-Frobenius theorem, see e.g. [2], describes some of the remarkable properties enjoyed by the eigenvalues and eigenvectors of irreducible nonnegative matrices (e.g. positive matrices).

Theorem 1. (Perron-Frobenius) Let A be an irreducible nonnegative square matrix. Then the spectral radius, $\rho(A)$, is a real eigenvalue, which has a positive (real) eigenvector. This eigenvalue called the principal eigenvalue of A is simple (it is not a multiple root of the characteristic equation), and its eigenvector called priority vector is unique up to a multiplicative constant.

The m-consistency of a nonnegative m-reciprocal $n \times n$ matrix A is measured by the *m-consistency index* $I_{mc}(A)$ defined in [6] as

$$I_{mc}(A) = \frac{\rho(A) - n}{n - 1}, \tag{14}$$

where $\rho(A)$ is the spectral radius of A (particularly, *the principal eigenvalue* of A).

Ranking the alternatives in X is determined by the vector of weights $w = (w_1, w_2, ..., w_n)$, with $w_i > 0$, for all $i = 1, 2, ..., n$, such that $\sum_{i=1}^{n} w_i = 1$, satisfying $Aw = \rho(A)w$, is called the (*normalized*) *principal eigenvector of* A, or, *priority vector of* A. Since the element of the priority vector w_i is interpreted as the relative importance of alternative x_i, the alternatives $x_1, x_2, ..., x_n$ in X are ranked by their relative importance. The following important result has been derived in [6].

Theorem 2. If $A = \{a_{ij}\}$ is an $n \times n$ positive m-reciprocal matrix, then $I_{mc}(A) \geq 0$. Moreover, A is m-consistent iff $I_{mc}(A) = 0$.

To provide a (in)consistency measure independently of the dimension of the matrix, n, T. Saaty in [6] proposed the consistency ratio. In order to distinguish it here from the other consistency measures, we shall call it *m-consistency ratio*. This is obtained by taking the ratio I_{mc} to its mean value R_{mc} over a large number of positive m-reciprocal matrices of dimension n, whose entries are randomly and uniformly generated, i.e.

$$CR_{mc} = \frac{I_{mc}}{R_{mc}}. \tag{15}$$

For this consistency measure it was proposed an estimation of 10% threshold of CR_{mc}. In other words, a pair-wise comparison matrix could be accepted (in a DM process) if its m-consistency ratio does not exceed 0.1, see [6]. The m-consistency index I_{mc} has been defined by (14) for m-reciprocal matrices, now, we shall investigate inconsistency/intransitivity also for a-reciprocal matrices. For this purpose we use relations between m-consistent and a-transitive/a-consistent matrices derived in Propositions 1 to 4. Let $B = \{b_{ij}\}$ be an a-reciprocal $n \times n$ matrix with $0 < b_{ij} < 1$ for all i and j. We define the *a-consistency index* $I_{ac}(B)$ of $B = \{b_{ij}\}$ as

$$I_{ac}(B) = I_{mc}(A), \text{where } A = \{\phi^{-1}(b_{ij})\}. \tag{16}$$

From (16) we obtain the following result which is parallel to Theorem 2.

Theorem 3. If $B = \{b_{ij}\}$ is an a-reciprocal $n \times n$ fuzzy matrix with $0 < b_{ij} < 1$ for all i and j, then $I_{ac}(B) \geq 0$. Moreover, B is a-consistent iff $I_{ac}(B) = 0$.

Now, we shall deal with measuring a-intransitivity of a-reciprocal matrices. Recall transformation functions φ_σ and φ_σ^{-1} defined by (10), (11), where $\sigma > 1$ is a given value. Let $B = \{b_{ij}\}$ be an a-reciprocal $n \times n$ matrix with $0 < b_{ij} < 1$ for all i and j. We define the *a-transitivity index* $I_{at}^{\sigma}(B)$ of $B = \{b_{ij}\}$ as

$$I_{at}^{\sigma}(B) = I_{mc}(A_\sigma), \text{where } A_\sigma = \{\varphi_\sigma^{-1}(b_{ij})\}. \tag{17}$$

From (11), (17) we obtain the following result which is parallel to Theorem 2 and 3.

Theorem 4. If $B = \{b_{ij}\}$ is an a-reciprocal $n \times n$ matrix with $0 < b_{ij} < 1$ for all i and j, then $I_{at}^{\sigma}(B) \geq 0$. Moreover, B is a-transitive iff $I_{at}^{\sigma}(B) = 0$.

Let $A = \{a_{ij}\}$ be an a-reciprocal $n \times n$ matrix. In (15), the *m-consistency ratio of* A denoted by $CR_{mc}(A)$ is obtained by taking the ratio $I_{mc}(A)$ to its mean value $R_{mc}(n)$ over a large number of randomly and uniformly generated positive m-reciprocal matrices of dimension n, i.e.

$$CR_{mc}(A) = \frac{I_{mc}(A)}{R_{mc}(n)}. \tag{18}$$

The table that gives the function values of $R_{mc}(n)$ can be found e.g. in [6]. Similarly, we define a-consistency ratio and a-transitivity ratio. Let $B = \{b_{ij}\}$ be an a-reciprocal $n \times n$ matrix with $0 < b_{ij} < 1$ for all i and j. We define the *a-consistency ratio* CR_{ac} of B as follows

$$CR_{ac}(B) = \frac{I_{ac}(B)}{R_{mc}(n)}. \tag{19}$$

The corresponding priority vector $w^{ac} = (w_1^{ac}, w_2^{ac}, ..., w_n^{ac})$ is given by the characteristic equation $\phi^{-1}(B)w^{ac} = \rho(\phi^{-1}(B))w^{ac}$.

Moreover, given $\sigma > 1$, we define *a-transitivity ratio* CR_{at}^{σ} of B as

$$CR_{at}^{\sigma}(B) = \frac{I_{at}^{\sigma}(B)}{R_{mc}(n)}. \tag{20}$$

The corresponding priority vector $w^{at} = (w_1^{at}, w_2^{at}, ..., w_n^{at})$ is given by $\varphi_\sigma^{-1}(B)w^{at} = \rho(\varphi_\sigma^{-1}(B)w^{at}$.

In practical DM situations a-inconsistency of a positive a-reciprocal pair-wise comparison matrix B is "acceptable" if $CR_{ac}(B) < 0.1$. Also, a-intransitivity of a positive a-reciprocal pair-wise comparison matrix B is "acceptable" if $CR_{at}^{\sigma}(B) < 0.1$. The final ranking of alternatives is given by the corresponding priority vector.

5 Illustrative example

Let $X = \{x_1, x_2, x_3, x_4\}$ be a set of 4 alternatives. The preferences on X are described by a positive matrix $B = \{b_{ij}\}$,

$$B = \begin{pmatrix} 0.5 & 0.6 & 0.6 & 0.9 \\ 0.4 & 0.5 & 0.6 & 0.7 \\ 0.4 & 0.4 & 0.5 & 0.5 \\ 0.1 & 0.3 & 0.5 & 0.5 \end{pmatrix}. \tag{21}$$

Here, $B = \{b_{ij}\}$ is a-reciprocal and it is a-inconsistent, as it may be directly verified by (7), e.g. $b_{12}.b_{23}.b_{31} \neq b_{32}.b_{21}.b_{13}$. At the same time, B is a-intransitive as $b_{12} + b_{23} + b_{31} = 1.9 \neq 1.5$. We consider $\sigma = 9$ and calculate

$$E = \{\phi^{-1}(b_{ij})\} = \begin{pmatrix} 1 & 1.50 & 1.50 & 9.00 \\ 0.67 & 1 & 1.5 & 2.33 \\ 0.67 & 0.67 & 1 & 1 \\ 0.11 & 0.43 & 1 & 1 \end{pmatrix},$$

$$F = \{\varphi_9^{-1}(b_{ij})\} = \begin{pmatrix} 1 & 1.55 & 1.55 & 5.80 \\ 0.64 & 1 & 1.55 & 2.41 \\ 0.64 & 0.64 & 1 & 1 \\ 0.17 & 0.42 & 1 & 1 \end{pmatrix}.$$

We calculate $\rho(E) = 4.29$, $\rho(F) = 4.15$. By (14), (19) and (20) we obtain $CR_{ac}(B) = 0.11 > 0.1$ with the priority vector $w^{ac} = (0.47; 0.25; 0.18; 0.10)$, which gives the ranking of alternatives $x_1 > x_2 > x_3 > x_4$. Similarly, $CR^9_{at}(B) = 0.056 < 0.1$ with the priority vector $w^{at} = (0.44; 0.27; 0.18; 0.12)$, giving the same ranking of alternatives $x_1 > x_2 > x_3 > x_4$.

As it is evident, a-consistency ratio $CR_{ac}(B)$ is too high that matrix B is considered a-consistent. On the other hand, a-transitivity ratio $CR^9_{at}(B)$ is sufficiently low that matrix B is considered a-transitive. The ranking of alternatives given by both methods remains, however, the same.

6 Conclusion

In this paper we investigated two types of pair-wise comparison matrices as well as the concepts of reciprocity, consistency/inconsistency and transitivity/intransitivity. In the literature, an inconsistency measure, i.e. inconsistency index, is known only for m-reciprocal matrix. Here we definedthe inconsistency index also for a-reciprocal matrices. As it was shown, the proposed concepts can be applied in ranking alternatives as well as in eliciting criteria weights in MCDM problems. New inconsistency/intransitivity indices will measure the quality of proposed ranking procedure. Numerical experiments show that there is no strong relationship between a-consistency and a-transitivity.

Acknowledgements

This research has been supported by GACR project No. 402090405.

References

[1] Chiclana, F. Herrera, F., Herrera-Viedma, E.: Integrating Multiplicative Preference Relations in a Multipur-pose Decision Making Model Based on Fuzzy Preference Relations. *Fuzzy Sets and Systems* 112 (2001) 277–291.

[2] Fiedler, M., Nedoma, J., Ramík, J., Rohn, J., Zimmermann, K.: *Linear Optimization Problems with Inexact Data*. Berlin - Heidelberg - New York - Hong Kong - London - Milan, Tokyo: Springer, 2006, 229 p.

[3] Fodor, J., Roubens, M.: *Fuzzy Preference Modelling and Multicriteria Decision Support*. Kluwer, Dordrecht, 1994.

[4] Ramík, J.: Pairwise comparison method in multi-criteria optimization - an alternative approach. In: *Proc. of the 29th International conference Mathematical Methods in Economics 2011*, September 6-9, Liptovsky Jan, Professional Publishing, Prague 2011, Part II., 587–592.

[5] Ramík, J., Vlach, M.: *Generalized concavity in optimization and decision making*. Kluwer Publ. Comp., Boston-Dordrecht-London, 2001, 305 p.

[6] Saaty, T. L.: *Multicriteria decision making - the Analytical Hierarchy Process*. Vol. I., RWS Publications, Pittsburgh, 1991, 479 p.

[7] Tanino, T.: Fuzzy Preference Relations in Group Decision Making. *Fuzzy Sets and Systems* 12 (1984) 117–131.

Fuzzy linear programming duality

Jaroslav Ramík [1], Milan Vlach[2]

Abstract. The word "duality" has been used in various areas of science for long time. Nevertheless, in general, there is a lack of consensus about the exact meaning of this important notion. However, in the field of optimization, and particularly in linear programming, the notion of duality is well understood and remarkably useful. Various attempts to develop analogous useful duality schemes for linear programming involving fuzzy data have been appearing since the early days of fuzzy sets. After recalling basic results on linear programming duality, we give examples of early attempts in extending duality to problems involving fuzzy data, and then we discuss recent results on duality in fuzzy linear programming and their possible application.

Keywords: linear programming, fuzzy linear programming, duality.

JEL classification: C61
AMS classification: 90C70

1 Introduction

As pointed out by Harold Kuhn [3] the elements of duality in optimization are: (i) A pair of optimization problem based on the same data, one a maximum problem with objective function $x \mapsto f(x)$ and the other a minimum problem with objective function $y \mapsto h(y)$. (ii) For feasible solutions x and y to the pair of problems, always $h(y) \geq f(x)$. (iii) Necessary and sufficient condition for optimality of feasible solutions $\bar{x}$ and $\bar{y}$ is $h(\bar{y}) = f(\bar{x})$.

This kind of duality is particularly clear, elegant, and remarkably useful in linear programming and its applications. Given the practical relevance of duality theory of linear programming, it is not surprising that attempts to develop analogous duality schemes for linear programming involving fuzzy data have been appearing since the early days of fuzzy sets [8]. To devise such a duality scheme, we have to specify in advance some class of permitted fuzzy numbers, define fundamental arithmetic operations with fuzzy numbers, and clarify the meaning of inequalities between fuzzy numbers. Because this can be done in inexhaustibly many ways, we can hardly expect a unique extension of duality to fuzzy situations, which would be so clean and clear like that of classical linear programming. Instead, there exist several variants of the duality theory for fuzzy linear programming, the results of which resemble in various degrees some of the useful results established in the conventional linear programming.

After recalling basic results of duality theory of linear programming, we first present early examples of pairs of mutually dual problems, in which only the inequalities $\leq$ and $\geq$ are allowed to become fuzzy. The feasible solutions of such problems are nonnegative vectors of a finite dimensional real vector space, and the degrees of constraints satisfaction and the degrees of optimality of feasible solutions are defined by the numerical data from the underlying linear programming problem and valued extensions of binary relations $\leq$ and $\geq$. Then we discuss duality pairs for problems in which some or all numerical data may also be fuzzy. The duality schemes for such problems are significantly more complicated because of necessity to extend $\leq$ and $\geq$ so that some consistent comparison of fuzzy quantities is possible. For reader's convenience of this extended abstract, we summarized necessary notions and results from the theory of fuzzy sets in the Appendix.

[1]Silesian University, Department of Mathematical Methods in Economics, Univerzitní náměstí 76, 733 40 Karviná Czech Republic, ramik@slu.cz

[2]Charles University, Department of Theoretical Informatics and Mathematical Logic, Malostranské náměstí 25, 118 00 Prague 1, Czech Republic, Milan.Vlach@mff.cuni.cz

2 Linear Programming Duality

Given real numbers $b_1, b_2, \ldots, b_m, c_1, c_2, \ldots, c_n, a_{11}, a_{12}, \ldots, a_{mn}$, we consider linear programming problems in the canonical form:

$$\begin{array}{lrcll} \text{Maximize} & c_1x_1 + c_2x_2 + \cdots + c_nx_n & & & (1) \\ \text{subject to} & a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n & \leq & b_i \quad i = 1, 2, \ldots, m & (2) \\ & x_j & \geq & 0 \quad j = 1, 2, \ldots, n & (3) \end{array}$$

Using the same data $b_1, b_2, \ldots, b_m, c_1, c_2, \ldots, c_n, a_{11}, a_{12}, \ldots, a_{mn}$, we construct another linear programming problem, called the *dual* problem to the *primal* problem (1)-(3), as follows:

$$\begin{array}{lrcll} \text{Minimize} & y_1b_1 + y_2b_2 + \cdots + y_mb_m & & & (4) \\ \text{subject to} & y_1a_{1j} + y_2a_{2j} + \cdots + y_ma_{mj} & \geq & c_j \quad j = 1, 2, \ldots, n & (5) \\ & y_i & \geq & 0 \quad i = 1, 2, \ldots, m & (6) \end{array}$$

It is easy to see that if one rewrites the dual problem into the form of the primal problem and again constructs the corresponding dual, then one obtains a linear programming problem which is equivalent to the original primal problem. In other words, the dual to the dual is the primal. Consequently, it is just the matter of convenience which of these problems is taken as the primal problem.

The well known results on the mutual relationships between the primal and the dual can be summarized as follows:

1. If x is a feasible solution of the primal problem and if y is a feasible solution of the dual problem, then $cx \leq yb$.
2. If $\bar{x}$ is a feasible solution of the primal problem, and if $\bar{y}$ is a feasible solution of the dual problem, and if $c\bar{x} = \bar{y}b$, then $\bar{x}$ is optimal for the primal problem and $\bar{y}$ is optimal for the dual problem.
3. If the feasible region of the primal problem is nonempty and the objective function $x \mapsto cx$ is not bounded above on it, then the feasible region of the dual problem is empty.
4. If the feasible region of the dual problem is nonempty and the objective function $y \mapsto yb$ is not bounded below on it, then the feasible region of the primal problem is empty.

It turns out that the following deeper results concerning mutual relation between the primal and dual problems hold:

5. If either of the problems (1)-(3) or (4)-(6) has an optimal solution, so does the other, and the corresponding values of the objective functions are equal.
6. If both of the problems (1)-(3) and (4)-(6) have feasible solutions, then both of them have optimal solutions and the corresponding optimal values are equal.
7. A necessary and sufficient condition that feasible solutions x and y of the primal and dual problems are optimal is that

$$\begin{array}{rcll} x_j > 0 & \Rightarrow & yA^j = c_j & 1 \leq j \leq n \\ x_j = 0 & \Leftarrow & yA^j > c_j & 1 \leq j \leq n \\ y_i > 0 & \Rightarrow & A_ix = b_i & 1 \leq i \leq m \\ y_i = 0 & \Leftarrow & A_ix < b_i & 1 \leq i \leq m \end{array}$$

where A^j and A_i stand for the j-th column and i-th row of $A = \{a_{ij}\}$, respectively.

It is also well known that the essential duality results of linear programming can be expressed as a saddle-point property of the Lagrangian function:

8. Let $\mathbb{R}^n_+$ and $\mathbb{R}^m_+$ denote the set of nonnegative n-vectors and m-vectors, and let $L\colon \mathbb{R}^n_+ \times \mathbb{R}^m_+ \to \mathbb{R}$ be the *Lagrangian function* for the primal problem (1)-(3), that is, $L(x,y) = cx + y(b - Ax)$. The necessary and sufficient condition that $\bar{x} \in \mathbb{R}^n_+$ be an optimal solution of the primal problem (1)-(3) and $\bar{y} \in \mathbb{R}^m_+$ be an optimal solution of the dual problem (4)-(6) is that $(\bar{x}, \bar{y})$ be a saddle point of L; that is, for all $x \in \mathbb{R}^n_+$ and $y \in \mathbb{R}^m_+$,

$$L(x, \bar{y}) \leq L(\bar{x}, \bar{y}) \leq L(\bar{x}, y) \tag{7}$$

3 Dual Pairs of Rödder and Zimmermann

One of the early approaches to duality in linear programing problems involving fuzziness is due to Rödder and Zimmermann [8]. To be able to state the problems considered by Rödder and Zimmermann concisely, we first observe that the conditions (7) bring up the pair of optimization problems

$$\text{maximize} \quad \min_{y\geq 0} L(x,y) \quad \text{subject to } x \in \mathbb{R}^n_+ \tag{8}$$
$$\text{minimize} \quad \max_{x\geq 0} L(x,y) \quad \text{subject to } y \in \mathbb{R}^m_+ \tag{9}$$

Let μ and μ' be real valued functions on $\mathbb{R}^n_+$ and $\mathbb{R}^m_+$, respectively; and let $\{\nu_x\}_{x\in\mathbb{R}^n_+}$ and $\{\nu'_y\}_{y\in\mathbb{R}^m_+}$ be families of real valued functions on $\mathbb{R}^m_+$ and $\mathbb{R}^n_+$, respectively. Furthermore, let φ_y and ψ_x be real valued functions on $\mathbb{R}^n_+$ and $\mathbb{R}^m_+$ defined by $\varphi_y(x) = \min(\mu(x), \nu_x(y))$ and $\psi_x(y) = \min(\mu'(y), \nu'_y(x))$. Now let us consider the following pair of families of optimization problems:

$$\begin{aligned} \text{Family } \{P_y\}: &\quad \text{Given } y \in \mathbb{R}^m_+, \text{ maximize } \ \varphi_y(x) \ \text{ subject to } \ x \in \mathbb{R}^n_+ \\ \text{Family } \{D_x\}: &\quad \text{Given } x \in \mathbb{R}^n_+, \text{ maximize } \ \psi_x(y) \ \text{ subject to } \ y \in \mathbb{R}^m_+ \end{aligned}$$

Motivated and supported by economic interpretation, Rödder and Zimmermann [8] propose to specify functions μ and μ' and families $\{\nu_x\}$ and $\{\nu'_y\}$ as follows: Given an $m \times n$ matrix A, $m \times 1$ vector b, $1 \times n$ vector c, and real numbers γ and δ, define the functions μ, μ', ν_x and ν'_y by

$$\mu(x) = \min(1, 1 - (\gamma - cx)), \quad \mu'(y) = \min(1, 1 - (yb - \delta)) \tag{10}$$
$$\nu_x(y) = \max(0, y(b - Ax)), \quad \nu'_y(x) = \max(0, (yA - c)x) \tag{11}$$

Strictly speaking, we do not obtain a duality scheme as conceived by Kuhn because there is no relationship between the numbers γ and δ. Indeed, if the the family $\{P_y\}_{y\geq 0}$ is considered to be the primal problem, then we have the situation in which the primal problem is completely specified by data A, b, c and γ. However, these data are not sufficient for specification of family $\{D_x\}_{x\geq 0}$ because the definition of $\{D_x\}_{x\geq 0}$ requires knowledge of δ. Thus from the point of view that the dual problem is to be constructed only on the basis of the primal problem data, every choice of δ determines a certain family dual to $\{P_y\}_{y\geq 0}$. In this sense we could say that every choice of δ gives a duality, the *δ-duality*. Analogously, if the primal problem is $\{D_x\}_{x\geq 0}$, then every choice of γ determines some family $\{P_y\}_{y\geq 0}$ dual to $\{D_x\}_{x\geq 0}$, and we obtain the *γ-duality*. In other words, for every γ, δ, we obtain (γ, δ)-duality. It is worth noticing that families $\{P_y\}$ and $\{D_x\}$ consist of uncountably many linear optimization problems. Moreover, every problem of each of these families may have uncountably many optimal solutions. Consequently, the solution of the problem given by family $\{P_y\}_{y\geq 0}$ is the family $\{X(y)\}_{y\geq 0}$ of subsets of $\mathbb{R}^n_+$ where $X(y)$ is the set of maximizers of φ_y over $\mathbb{R}^n_+$. Analogously, the family $\{Y(x)\}_{x\geq 0}$ of maximizers of ψ_x over $\mathbb{R}^m_+$ is the solution of problem given by family $\{D_x\}_{x\geq 0}$. Rödder and Zimmermann propose to replace the families $\{P_y\}$ and $\{D_x\}$ by the families $\{P'_y\}$ and $\{D'_x\}$ of problems defined as follows:

$$\text{maximize} \quad \lambda \quad \text{subject to } \lambda \leq 1 + cx - \gamma, \ \lambda \leq y(b - Ax), \ x \geq 0 \tag{12}$$
$$\text{minimize} \quad \eta \quad \text{subject to } \eta \geq yb - \delta - 1, \ \eta \geq (c - yA)x, \ y \geq 0 \tag{13}$$

They call these families of optimization problems the *fuzzy dual pair* and claim that the families $\{P_y\}$ and $\{D_x\}$ become families $\{P'_y\}$ and $\{D'_x\}$ when μ, μ', ν_x and ν'_y are defined by (10)-(11). To see that this claim cannot be substantiated, it suffices to observe that the value of function φ_y cannot be greater than 1, whereas the value of λ is not bounded above whenever A and b are such that both cx and $-yAx$ are positive for some $x \in R^n_+$. To obtain a valid conversion, one needs to add the inequalities $\lambda \leq 1$ and $\eta \geq -1$ to the constraints. Thus it seems that more suitable choice of functions ν_x and ν'_y in the Rödder

and Zimmermann duality scheme would be $\nu_x(y) = \min(1, 1+y(b-Ax))$ and $\nu'_y(x) = \min(1, 1+(yA-c)x)$. Another objection to the Rödder and Zimmermann model arises from the fact that, the duality results for the proposed fuzzy dual pair do not reduce to the standard duality results for the crisp scenario, that is, for $\lambda = 1, \eta = -1$. Again an easy remedy is to work with $\nu_x(y) = \min(1, 1 + y(b - Ax))$ and $\nu'_y(x) = \min(1, 1 + (yA - c)x)$ instead of $\nu_x(y) = \min(0, 1 + y(b - Ax))$ and $\nu'_y(x) = \min(0, 1 + (yA - c)x)$.

4 Duality Pairs of Bector and Chandra

In contrast to the usual practice, in the Rödder and Zimmermann model, the range of membership functions μ and μ' is $(-\infty, 1]$, and the range of membership functions ν_x and ν'_u is $[0, \infty)$ or $[1, \infty)$ instead of usual $[0, 1]$. Bector and Chandra [1] propose to replace the relations $\leq$ and $\geq$ appearing in the dual pair of linear programming problems by their valued extensions. In particular, the inequality $\leq$ appearing in the ith constraint of the primal problem is replaced by its valued extension $\preceq_i$ whose membership function $\mu_{\preceq_i}$: $\mathbb{R} \times \mathbb{R} \to [0, 1]$ is defined by

$$\mu_{\preceq_i}(\alpha, \beta) = \begin{cases} 1 & \text{if} \quad \alpha \leq \beta \\ 1 - \frac{\alpha - \beta}{p_i} & \text{if} \quad \beta < \alpha \leq \beta + p_i \\ 0 & \text{if} \quad \beta + p_i < \alpha \end{cases}$$

where p_i is a positive number. Analogously, the inequality $\geq$ appearing in the jth constraint of the dual problem is replaced by its valued relation $\succeq_j$ with membership function

$$\mu_{\succeq_j}(\alpha, \beta) = \begin{cases} 1 & \text{if} \quad \alpha \geq \beta \\ 1 - \frac{\beta - \alpha}{q_j} & \text{if} \quad \beta > \alpha \geq \beta - q_j \\ 0 & \text{if} \quad \beta - q_j > \alpha \end{cases}$$

where q_j is a positive number. The degree of satisfaction with which $x \in \mathbb{R}^n$ fulfills the ith fuzzy constraint $A_i x \preceq_i b_i$ of the primal problem is expressed by the fuzzy subset of $\mathbb{R}^n$ whose membership function μ_i is defined by $\mu_i(x) = \mu_{\preceq_i}(A_i x, b_i)$, and the degree of satisfaction with which $y \in \mathbb{R}^m$ fulfills the jth fuzzy constraint $yA^j \succeq_j c_j$ of the dual problem is expressed by the fuzzy subset of $\mathbb{R}^m$ whose membership function μ_j is defined by $\mu_j(y) = \mu_{\succeq_j}(yA^j, c_j)$. Similarly, we can express the degree of satisfaction with a prescribed aspiration level γ of the objective function value cx by the fuzzy subset of $\mathbb{R}^n$ given by $\mu_0(x) = \mu_{\succeq_0}(cx, \gamma)$ where, for the tolerance given by a positive number p_0, the membership function $\mu_{\succeq_0}$ is defined by

$$\mu_{\succeq_0}(\alpha, \beta) = \begin{cases} 1 & \text{if} \quad \alpha \geq \beta \\ 1 - \frac{\beta - \alpha}{p_0} & \text{if} \quad \beta > \alpha \geq \beta - p_0 \\ 0 & \text{if} \quad \beta - p_0 > \alpha \end{cases}$$

Analogously, for the degree of satisfaction with the aspiration level δ and tolerance q_0 in the dual problem, we have $\mu_0(y) = \mu_{\preceq_0}(\delta, yb)$ where

$$\mu_{\preceq_0}(\alpha, \beta) = \begin{cases} 1 & \text{if} \quad \alpha \leq \beta \\ 1 - \frac{\alpha - \beta}{q_0} & \text{if} \quad \beta < \alpha \leq \beta + q_0 \\ 0 & \text{if} \quad \beta + q_0 < \alpha \end{cases}$$

This leads to the following pair of linear programming problems:

Given positive numbers $p_0, p_1, \ldots, p_m$, and a real number γ, maximize λ subject to

$$\begin{aligned} &(\lambda - 1)p_0 \leq cx - \gamma \\ &(\lambda - 1)p_i \leq b_i - A_i x, \ 1 \leq i \leq m \\ &0 \leq \lambda \leq 1, x \geq 0 \end{aligned} \tag{14}$$

Given positive numbers $q_0, q_1, \ldots, q_n$, and a real number δ, minimize $-\eta$ subject to

$$\begin{aligned} &(\eta - 1)q_0 \leq \delta - yb \\ &(\eta - 1)q_j \leq yA^j - c_j, \ 1 \leq j \leq n \\ &0 \leq \eta \leq 1, y \geq 0 \end{aligned} \tag{15}$$

Bector and Chandra call this pair the *modified fuzzy pair* of primal dual linear programming problems. Again we see that the dual problem is not stated by using only the data available in the primal problem. Indeed, if the problem (14) is considered to be the primal problem, then to state its dual problem one needs additional information; namely, a number δ and numbers $q_0, q_1, \ldots, q_n$; if problem (15) is considered to be primal, then one needs a number γ and numbers $p_0, p_1, \ldots, p_m$.

5 Duality Pairs of Ramík

As mentioned in the Introduction, when we wish to develop a sensible duality scheme for linear programing problems in which also the numerical data may be fuzzy, then we need tools for comparing fuzzy numbers. Recently, Ramík [6] and [5] (see also [2]) proposed a rather general duality scheme in which the fuzzy quantities are compared by means of extensions of binary relations $\leq$ and $\geq$ on $\mathbb{R}$ to fuzzy relations on $\mathcal{F}(\mathbb{R}) \times \mathcal{F}(\mathbb{R})$. Moreover, this scheme does not require external specification of goals or aspiration levels, and a number of earlier duality schemes can be obtained as special cases. A simple version of this scheme can briefly be described as follows.

Given fuzzy numbers $B_1, B_2, \ldots, B_m; C_1, C_2, \ldots, C_n; A_{11}, A_{12}, \ldots, A_{mn}$ from some class of fuzzy numbers, and fuzzy extensions $\preceq_1, \ldots, \preceq_m; \succeq_1, \ldots, \succeq_n$ of $\leq$ and $\geq$, respectively, we construct the pair of problems

$$\begin{aligned}
&\text{Maximize} && C_1x_1 + C_2x_2 + \cdots + C_nx_n && && && (16)\\
&\text{subject to} && A_{i1}x_1 + A_{i2}x_2 + \cdots + A_{in}x_n && \preceq_i && B_i \qquad i = 1, 2, \ldots, m && (17)\\
& && x_j && \geq && 0 \qquad j = 1, 2, \ldots, n && (18)
\end{aligned}$$

$$\begin{aligned}
&\text{Minimize} && y_1B_1 + y_2B_2 + \cdots + y_mB_m && && && (19)\\
&\text{subject to} && y_1A_{1j} + y_2A_{2j} + \cdots + y_mA_{mj} && \succeq_j && C_j \qquad j = 1, 2, \ldots, n && (20)\\
& && y_i && \geq && 0 \qquad i = 1, 2, \ldots, m && (21)
\end{aligned}$$

where "+" is defined by the standard extension principle, and where the meanings of "feasibility" and "optimality" are specified as follows.

Let β be a positive number from $[0, 1]$. By a *β-feasible region* of problem (16)-(18) we understand the β-cut of fuzzy subset X of $\mathbb{R}^n$ given by membership function

$$\mu_X(x) = \begin{cases} \min\limits_{1 \leq i \leq m} \mu_{\preceq_i}(A_{i1}x_1 + \cdots + A_{in}x_n, B_i) & \text{if } x_j \geq 0 \text{ for all } j \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

and by a *β-feasible solutions* of problem (16)-(18) we understand the elements of *β-feasible region.*

To explain "maximization", we first observe that a feasible solution $\bar{x}$ of non fuzzy problem (1)-(3) is optimal exactly when there is no feasible solution x such that $cx > c\bar{x}$. This suggests to consider a fuzzy extension $\succeq$ of $\geq$ and to introduce, for each positive α from $[0, 1]$, the binary relations $\geq_\alpha$ and $<_\alpha$ on $\mathcal{F}(\mathbb{R})$ by "$a \geq_\alpha b$" means $\mu_{\succeq}(a, b) \geq \alpha$, and "$a <_\alpha b$" means ($\mu_{\succeq}(a, b) \geq \alpha$ and $\mu_{\succeq}(b, a) < \alpha$). Now let α and β be positive numbers from $[0, 1]$. We say that a β-feasible solution $\bar{x}$ of (16)-(18) is (α, β)-*maximal solution* of (16)-(18) if there is no β-feasible solution of (16)-(18) x different from $\bar{x}$ such that $C_1\bar{x}_1 + C_2\bar{x}_2 + \cdots + C_n\bar{x}_n <_\alpha C_1x_1 + C_2x_2 + \cdots + C_nx_n$. The notions of β-feasibility and (α, β)-minimality for the dual problem (19)-(21) are defined analogously.

6 Appendix

Let $\mathcal{U}$ be a fixed nonempty set and let X, Y and Z be subsets of $\mathcal{U}$. Recall that if f is a function from Y to Z and X is a subset of Y, then the function g from X to Z such that $g(x) = f(x)$ for all $x \in X$ is called the *restriction* of f to X, and f is called an *extension* of g to Y. If A is a subset of X, then the *characteristic function* of A is the function χ_A from X to $\{0, 1\}$ such that $\chi_A(x) = 1$ for $x \in X$ and $\chi_A(x) = 0$ otherwise.

The phrase "membership function of a fuzzy set" is very common one in the fuzzy set literature. Obviously such a phrase and similar ones strongly suggest that "fuzzy sets" and "membership functions

of fuzzy sets" are different objects. We follow the opinion that fuzzy sets are special nested families of subsets of a set. In more detail (for full details, see [4] or [7]), a *fuzzy subset* A of X is the family $\{A_\alpha\}_{\alpha\in[0,1]}$ of subsets of X such that $A_0 = X, A_\beta \subset A_\alpha$ whenever $0 \leq \alpha \leq \beta \leq 1$, and $A_\beta = \cap_{0\leq\alpha<\beta} A_\alpha$ whenever $0 < \beta \leq 1$. If A is a fuzzy subset of X, then the *membership function of* A is the function $\mu_A : X \to [0,1]$ defined by $\mu_A(x) = \sup\{\alpha : x \in A_\alpha\}$, and the function value $\mu_A(x)$ is called the *membership degree* of x in A. For each $\alpha \in [0,1]$, the set $\{x \in X \mid \mu_A(x) \geq \alpha\}$ is called the *α-cut of A*. It is worth noting that if f is an arbitrary function from X into $[0,1]$, then the family $A = \{A_\alpha\}_{\alpha\in[0,1]}$ of sets $\{x \in X \mid f(x) \geq \alpha\}$ is a fuzzy subset of X and f is a membership function of A. Moreover, if μ_A is the membership function of a fuzzy subset of A, then the α-cut of A coincides with A_α for each $\alpha \in [0,1]$. Therefore, there is a natural one-to-one correspondence between fuzzy subsets of X and real functions from X to $[0,1]$, and each fuzzy subset A of X can be specified by its membership function μ_A and vice-versa. Consequently, it does not matter whether we introduce or discuss the properties of fuzzy subsets of a set in terms of families subsets or membership functions, and the meaning of phrases like "the fuzzy subset of X determined by a membership function $\mu : X \to [0,1]$" or "the fuzzy set $\mu : X \to [0,1]$" becomes clear. Because of the existence of one-to-one correspondence between the subsets and characteristic functions of subsets and because there is also a one-to-one correspondence between the characteristic functions of subsets and the membership functions with values in $\{0,1\}$, we can view subsets of X as fuzzy subsets of X. When we need to distinguish the latter from the other fuzzy subsets of X, we call them the *crisp fuzzy subsets* of X. We denote the collection of all fuzzy subsets of X by $\mathcal{F}(X)$ and the collection of all crisp fuzzy subsets of X by $\mathcal{P}(X)$.

If A is from $\mathcal{F}(X)$, then the set $\{x \in X : \mu_A(x) = 1\}$ is called the *core* of A. If B is from $\mathcal{F}(X)$ and A is from $\mathcal{P}(X)$, and if $\mu_B(x) = \mu_A(x)$ for all x in the core of A, then we say that B is a *valued extension* of A. Because of the existence of one-to-one correspondence between subsets of X and the elements of $\mathcal{P}(X)$, we also have *valued extensions of subsets* of X. Recall that subsets of $X \times X$ are called *binary relations* on X and fuzzy subsets of $X \times X$ are called *fuzzy relation* on X. Applying the previous construction to $X \times X$, we obtain valued extensions of crisp fuzzy relations on X and valued extensions of binary relations on X. Finally, by similar construction we can obtain a valued extensions of fuzzy relations on X to fuzzy relations on $\mathcal{F}(X)$.

Acknowledgements

This research has been supported by GACR project No. 402090405

References

[1] Bector, C. R., and Chandra, C.: On duality in linear programming under fuzzy environment, *Fuzzy Sets and Systems* **125** (2002), 317–325.

[2] Inuiguchi, M., Ramík, J., Tanino, T., and Vlach, M.: Satisficing solutions and duality in interval and fuzzy linear programming, *Fuzzy sets and systems* **135** (2003), 151–177.

[3] Kuhn, H. W.: Nonlinear Programming: A Historical View. In: *SIAM–AMS Proceedings, Volume 9*, AMS, Providence, Rhode Island, 1976, 1–26.

[4] D. A. Ralescu, D. A.: A generalization of the representation theorem, *Fuzzy Sets and Systems* **51** (1992), 309–311.

[5] Ramík, J.: Duality in fuzzy linear programming: Some new concepts and results, *Fuzzy Optimization and Decision Making* **4** (2005), 25–39.

[6] Ramík, J.: Duality in fuzzy linear programming with possibility and necessity relations, *Fuzzy Sets and Systems* **157** (2006), 1283–1302.

[7] Ramík, J., and Vlach, M.: A non-controversial definition of fuzzy sets. In *Transactions on rough sets II - Rough sets and fuzzy sets.* (J. F. Peters, and A. Skowron, eds.), Springer-Verlag, Berlin, Heidelberg, 2004, 201–207.

[8] Rödder, W., and Zimmermann, H-J.: Duality in fuzzy linear programming. In: *Extremal Methods and System Analysis* (A. V. Fiacco, and K. O. Kortanek, eds.), Berlin, New York, 1980, 415–429.

Consumption in the Baltic States: Myopia or liquidity constraints?

Svetlana Ridala[1], Ants Aasma[2]

Abstract. This paper investigates the consumption pattern of the Baltic States for the period 1999-2010. The investigation is done using the aggregate annual data from the Baltic States: Estonia, Latvia and Lithuania. The panel data analysis of the data rejects the life-cycle - permanent-income hypothesis (LCH - PIH): the fraction of income, attributed to the "rule of thumb" consumption is generally substantial for these countries. The paper checks the possible reasons for this result with the well-known Shea's model with three hypotheses: myopia, liquidity constraints and "opposite asymmetry". Additionally the results based on the data of consumption of non-durables and services are compared with the results based on the data of total consumption.

Keywords: "rule-of-thumb" consumption, liquidity constraints, myopia, "opposite asymmetry".

JEL Classification: E21, C22
AMS Classification: 91B42, 91B84, 62M10, 62P20

1 Introduction

In recent years, there has been an increasing interest in consumer choice in economics. This problem has been analyzed mainly with two models: overlapping generation's model (Diamond [6]) and permanent-income model (Friedman [8], Hall [11]). During the last 25 years several papers have been published that do not support the results of these models. The most central among above-mentioned works are the papers of Campbell and Mankiw [2]-[4]. They considered that the world is populated with two types of consumers: a) individuals who consume their current income ("rule-of-thumb" consumers), the share of income, accrued to such consumers, is denoted by λ; b) individuals who consume their permanent income, i.e. they use their extra income for saving and follow life cycle-permanent income hypothesis (LCH-PIH); the share of income, accrued to such individuals, is $1-\lambda$. Using aggregate post-war US quarterly time-series data from 1948:1 through 1985:1 Campbell and Mankiw estimated that half of the consumers do not save the income gained from tax cuts, but spend it on consumption. Later on the LCH-PIH was rejected by several authors (for example, Chyi and Huang [5], Filer and Fisher [7], Fuhrer [9], Gomes and Paz [10], Sarantis and Stewart [19], Shea [20], Souleles [22]). The reasons vary. Households with less income are liquidity-constrained and just don't have the funds for saving (e.g., Souleles [22]). Some families may also be too optimistic about the future, hoping that their income and level of welfare increases. Several authors have found that the "rule-of-thumb" behavior is an essential cause for the rejection of LCH-PIH. Other reasons include habit persistence (e.g. Fuhrer [9]) and non-separable preferences between consumption and leisure/labor supply (e.g., Kiley [13]).

This article takes a country-specific approach in analyzing the relationship between consumption and disposable income in the Baltic States. The Baltic countries, Estonia, Latvia and Lithuania form the homogenous group for analyzing economical models and share many similarities like location, similar size and historic background. In this paper the investigations, started in [17], were continued. In [17] the LCH-PIH hypothesis was tested for the Baltic States for the period 1996-2010 using GMM estimation technique on cross country time series of aggregate quarterly data. Both in [17] and in the present paper the approach of Campbell and Mankiw with the two groups of consumers are used. During the period 1996-2010 the countries experienced rapid growth (1999 - 2007) and a recession (2007 – 2010). For both periods the LCH-PIH hypothesis is rejected: about 40-90% of income was consumed by current income consumers. In this paper the panel data techniques and aggregate annual data for the period of 1999 – 2010 were used for testing the LCH-PIH. It was established that the fraction of income, attributed to the "rule of thumb" consumers, is generally substantial for these countries: 81% of the total consumption and 74% of consumption of non-durables and services follow the "rule of thumb" consumption. These estimates are in good accordance with the results of Kukk et al. [14], who, using Estonian micro-data for

[1] Tallinn University of Technology / Tallinn School of Economics and Business Administration, Department of Economics, Akadeemia tee 3, Tallinn 12618, e-mail: Svetlana.Ridala@tseba.ttu.ee.
[2] Tallinn University of Technology / Tallinn School of Economics and Business Administration, Department of Economics, Akadeemia tee 3, Tallinn 12618, e-mail: Ants.Aasma@tseba.ttu.ee.

the period 2002 – 2007, estimated that the fraction of "rule of thumb" consumption was approximately 75% in this period for Estonia. The possible reasons behind such estimates are checked using the Shea's [21] model with three hypothesis: myopia, liquidity constraints and "opposite asymmetry". Following Sarantis and Steward [19] (who suggested that the λ-model could allow "rule-of-thumb" consumers to have a durable consumption component), the estimates, based on the data of consumption of non-durables and services, are compared with the results, based on the data of total consumption.

The paper is organized as follows. Section 2 presents an overview of the related literature that gives different empirical estimates of the share of "rule-of-thumb" consumption from total income. The third section presents the data, methodology and empirical results. The fourth contains the conclusion.

2 Literature review

This section gives a little overview of different works, where LCH-PIH was rejected. Mainly we concentrate to the papers, where different estimates of the fraction of income, accrued to "rule-of-thumb" consumption, are presented.

One of the most important works dedicated to the problem of the estimation of λ is the paper of Weber [24], where the comparisons of the previous estimates of λ by different authors are provided. He argues that most authors have used log-linear Euler equation:

$$\Delta c_{t+1} = \mu + \Delta y_{t+1} + \sum_{i=1}^{k} \gamma_i X_i + \varepsilon_{t+1} , \tag{1}$$

where t is the time period, Δc_t is the change in log consumption spending, Δy_t is the change in log disposable income, the X_i's are other variables included in regression, such as real asset returns, and the γ_i's are slope coefficients, μ is the intercept and ε_t is the random error. Different researches estimated the value of λ for different countries and for different periods and found that the values of λ are between 0,114 and 0,843.

Several authors have rejected the LCH-PIH using data for the US. Fuhrer [9] estimated that the share of "rule-of-thumb" consumption is 29% by the Generalized Method of Moments (GMM), and 26% by the Full Information Maximum Likelihood (FIML). He also found that very important evidence against LCH-PIH is provided by the habit persistence. Additional evidence against LCH-PIH was also provided by Shea [20] ("opposite asymmetry"), Souleles [22] (liquidity-constrained consumers) and Kiley [13] (non-separable preferences between consumption and leisure/labor supply). Very recently Sahm et al. [18] analyzed fiscal stimulus packages of Michigan from 2008 and 2009 and found that 25% of households increased spending due to spring 2008 tax rebate and 13% due to spring 2009 tax rebate. One explanation of the variation of the "rule-of-thumb" consumption is recession. The share of "rule-of-thumb" consumption may be higher due to higher credit constraints or lower due to expectations of lower life-time income.

The estimates of λ indicate differences across countries in the effect of disposable income and consumption: for the United Kingdom with the sample period 1957-1988 (seasonally adjusted) λ = 0,203, for Canada with the period 1972 – 1988 (seasonally adjusted) λ = 0,225, for France with the sample period 1972 – 1988 (seasonally adjusted) λ = 0,401, for Japan with period 1972 – 1988 (not seasonally adjusted) λ = 0,035 and for Sweden with the period 1972 – 1988 (not seasonally adjusted) λ = 0,357 (Campbell and Mankiw [4]). For Japan, using a different sample period of 1957 – 1990 and different instrument set, Chyi and Huang [5] got another estimate: λ = 0,685. They also got the estimates of λ for some East Asian countries: λ = 0,333 for Korea with the annual sample period of 1966 – 1989, λ = 0,443 for the Philippines with the period of 1970 – 1990, λ = 0,412 for Thailand with the same period and λ = 0,275 for Taiwan with the sample period of 1961 – 1990.

Sarantis and Stewart [19] found that the presence of "rule-of-thumb" consumers is the major factor for the rejection of the basic LCH-PIH model in all OECD countries, which is firstly due to the liquidity constraints and a lesser extent due to the precautionary saving. The share of current income consumption varies across countries with the minimum of 44.2% for the UK and the maximum of 94.4% for Finland. The average proportion of income, accrued to the current income consumers, was calculated to be approximately 71% from time series estimates and 67% from panel estimates. Real per-capita total consumer expenditure was used in the estimates.

Gomes and Paz [10] checked the aggregate consumption behavior in four South American countries: Brazil, Colombia, Peru and Venezuela. They found that the estimation $\lambda \in [0.828, 0.906]$ for Columbia is statistically significant at 5% level for different instruments, for Brazil: $\lambda \in [0.734, 1.063]$ is statistically significant at 5% level, while the estimation $\lambda = 1.063$ had a big standard error 0,318, for Venezuela: $\lambda \in [0.703, 1.043]$ is statistically significant at 5% level. They noted that similar results for Brazil were also presented by other authors like Reis et al. [16]: $\lambda = 0.8$, Issler and Rocha [12]: $\lambda = 0.74$. Gomes and Paz displayed very surprising results for

Colombia: $\lambda \in [1.245, 1.502]$ for all instrument lists at 5 % level of statistical significance. All estimates of λ are larger than one that is not consistent with Campbell's and Mankiw's approach. They suggested that the total consumption has a specification problem, or alternatively, the predictive power of instruments is low.

3 Econometric analysis

3.1 Data

In this research aggregated annual data from the period 1999 – 2010 are used. Real adjusted gross disposable income of households per capita (*Y*) in PPPs was downloaded from Eurostat. The real total consumption (*TC*) is calculated as the adjusted final consumption expenditure of households divided by the purchasing power parities (PPPs) of the actual individual consumption of households and by the total resident population; all time-series obtained from the Eurostat. The consumption of non-durables and services were obtained from the OECD's database for Estonia. For Latvia and Lithuania the data was obtained from Latvian and Lithuanian national statistics authorities. The data was converted into the real terms using PPPs. The real interest rate is the lending interest rate adjusted for inflation as measured by the GDP deflator. This was downloaded from the World Bank database.

For Estonia, the nominal interest rates on one-year time deposits were obtained from the Bank of Estonia. For Latvia and Lithuania, one-year nominal interest rates were downloaded from DataStream.

3.2 The model

Following Campbell and Mankiw [2], the total change on consumption C_t (non-durables and services) with the assumption that real interest rate is not constant can be presented as follows:

$$\Delta \ln C_t = \mu + \lambda \Delta \ln Y_t + \phi r_t + \xi_t \,, \tag{2}$$

where $\Delta \ln C_t$ is the change of logarithm consumption spending and $\Delta \ln Y_t$ is the change of logarithm disposable income Y_t, r_t is the real interest rate, μ is the intercept, ξ_t is the random error, ϕ is the regression parameter for r_t. It is assumed that ξ_t is not correlated with the information from the period $t - 1$, but may be correlated with r_t and ΔY_t. Under LCH-PIH λ should be equal to zero.

Following Shea [21] the presence of myopia or liquidity constraints can be tested by the equation

$$\Delta \ln C_t = \mu + \lambda_1 (Pos_t) \Delta \ln Y_t + \lambda_2 (Neg_t) \Delta \ln Y_t + \phi r_t + \xi_t \,, \tag{3}$$

where Pos_t is a dummy variable equal to 1 for periods in which $\Delta \ln Y_t > 0$ and zero otherwise, and $Neg_t = 1 - Pos_t$, i.e. is equal to 1 when $\Delta \ln Y_t < 0$. The LCH-PIH implies that $\lambda_1 = \lambda_2 = 0$. Under myopia, λ_1 and λ_2 should be equal, and significantly larger than zero. Under liquidity constraints the condition $\lambda_1 > \lambda_2 > 0$ must be true at statistically significant level. The opposite case, when consumption is more sensitive to decrease rather to increase, i.e. $\lambda_1 < \lambda_2$, presents "opposite asymmetry".

Total consumption, TC_t, may be expressed as: $TC_t = C_t^{\eta_1} (CD_t)^{\eta_2}$, where C_t and CD_t denote correspondingly non-durable and durable expenditures and η_1, η_2 are positive constants. We suppose that the innovations to durable (v_t) and non-durable (ε_t) are proportional, i.e. $v_t / \eta_1 \varepsilon_t \equiv \rho$. Then, following Sarantis and Stewart [19], Gomes and Paz [10], and equations (2), (3), the total consumption can be estimated with the help of the following equations:

$$\Delta \ln TC_t = \mu + \lambda \Delta \ln Y_t + \phi r_t + \tau_t + \xi_1 \tau_{t-1} + \xi_2 \tau_{t-2} \tag{4}$$

$$\Delta \ln TC_t = \mu + \lambda_1 (Pos_t) \Delta \ln Y_t + \lambda_2 (Neg_t) \Delta \ln Y_t + \phi r_t + \tau_t + \xi_1 \tau_{t-1} + \xi_2 \tau_{t-2} \,, \tag{5}$$

where $\tau_t = (1 - \lambda)\eta_1 \varepsilon_t$, $\xi_1 = \rho \eta_2$ and $\xi_2 = A \rho \eta_2$ (A is a constant).

3.3 Results

For the estimation of λ we use the panel data approach [1] since it is not possible to realize the usual time series analysis due to very short time period: 12 years. We have annual data only for the period 1999 – 2010; earlier data for those countries are not available. As we have data for three countries for the period 1999 – 2010, we can use 36 observations in total. To determine the stationarity of variables, the empirical analysis starts from the panel unit root test. Namely, we apply the Levin-Lin-Chu test (Table 1), because this test imposes homogeneity on the unit root coefficient, while, for example, Im-Pesaran-Shin test does not. If panels are non-stationary, then any conclusion of the panel estimates will be invalid due to inconsistency and falseness. The test shows that the data are stationary in level I(0) for all variables.

Variable	Level		1st Difference		Process
	t-value	*P* > *t*	*t*-value	*P* > *t*	
Y	-1.600	0.0548	-1.669	0.0476	I(0)
	(-2.301)		(-3.882)		
TC	-1.689	0.0456	-6.288	0.0000	I(0)
	(-2.887)		(-8.077)		
C	-1.772	0.0382	-3.778	0.0001	I(0)
	(-2.810)		(-5.109)		
i	-2.076	0.0190	-2.881	0.0020	I(0)
	(-2.169)		(-2.985)		
r	-2.703	0.0034	-4.502	0.0000	I(0)
	(-2.818)		(-4.704)		

Table 1 Levin-Lin-Chu stationarity test

Note: In the analysis, Akaike criteria's automatic selection of lags was used; automatic Newey-West bandwidth selection method (with Bartlett kernel type) was used, p – statistic was calculated using normal distribution.

The results of the estimates of equations (2) and (4) are presented in Table 2. The first and the third column show the estimates for the total consumption, and the second and the forth show the estimations for consumption of non-durables and services. The OLS (ordinary least squares) and Arellano and Bond GMM (general method of moments) estimation techniques are used. Equations (4) and (5) have a MA(2) error process which requires a special care in choosing the instruments, namely the first two lags should be excluded from the estimations. The expanding GMM instruments – the real and nominal interest rates, change of consumption and disposable income - are lagged 2-3 years similarly to Sarantis and Stewart [19]. In order to choose between fixed effects (FE) and random effects (RE) model (in accordance with [23]) we applied the Hausman test (the results not presented). The test suggested that the RE approach is more appropriate. Although the OLS estimation is biased (see, for example [3]), we presented the OLS estimates for comparison with GMM estimates.

	TC OLS	*C* OLS	*TC* GMM	*C* GMM
λ	0.6719***	0.6219***	0.8115***	0.7365***
	(0.1586)	(0.1538)	(0.2843)	(0.2724)
J	-	-	6.3717	7.3904
Jp	-	-	[0.497]	[0.389]
R-squared	0.3911	0.3772	0.369	0.343

Table 2 Estimates of equations (2) and (4) in period 1999 – 2010

*Note: μ and ϕ included, not shown. The expanding GMM instruments are lagged 2-3 years. The test of restrictions is Sargan's test of overidentifying restrictions (H_0: instruments are valid) with J-statistic and p-value in square brackets with the chi-square distribution. Robust standard errors are in parenthesis. ***, ** and * denote statistical significance at 1%, 5% and 10% respectively.*

From Table 2 we see that OLS and GMM estimates are different, but this difference is not very great. More trustworthy estimates we got with the GMM method. So it can be asserted that the bulk of the Baltic consumers follow the "rule-of-thumb" behavior, they make 81% of the total consumption and 74% of the consumption of non-durables and services (Table 2). Only a small part of the consumption follows the PIH-LCD approach: 19% and 26% respectively. As it can be seen from the estimates, the difference between consumption of non-durables and services and total consumption is present, but it is not so big as to make the total consumption model inconsistent. The model of Sarantis and Stewart [19] with the allowance of durable component can be used in the first

instance for the countries, where the data for durability are not available. However, we must admit that the estimates of λ contain certain ambiguity, because we have very short time-series for trustworthy estimation.

Due to the above-mentioned estimates of λ the LCD-PIH is inconsistent for the Baltic countries. To find the possible reason for this the Shea's model (3) for consumption of non-durables and services and modified. Shea's model (5) for consumption with the durable component was used. OLS and GMM with the same instruments like in previous estimations were used and are presented in Table 3. The coefficient of positive change of disposable income λ_1 is positive and smaller than the coefficient of negative change for disposable income λ_2. It indicates the "opposite asymmetry": the aggregate consumption is more strongly correlated with the decrease of predictable income than increase. However, the estimates of λ_1 for consumption of non-durables and services become insignificant, weakening so the conclusion about the "opposite asymmetry". As the F-test resulted with $\lambda_1 \neq \lambda_2$, then the myopic behavior is rejected. The value of λ_2 is bigger than one: $\lambda_2 = 1.4684$ with the standard error 0.4158 for *TC* GMM and $\lambda_2 = 1.3975$ with the standard error 0.3931 for *C* GMM. These results are in contradiction with the rule that the elasticity of consumption with respect to disposable income should be less or equal to one. It should be mentioned that similar estimates for λ_2 are also presented by Shea [21] and for λ_1 (for Colombia) by Gomes and Paz [10] (moreover, for λ_2 their estimates are insignificant for all examined cases). However, the estimates, presented here, are just a little better: taking into account the standard error, it can be asserted that with a little probability the real value of λ_2 could be approximately equal 1, but Gomes and Paz and Shea estimated that the real values of λ_1 and λ_2 are firmly greater than 1. Similarly to [10] and [21] our estimates for λ_2 may be caused by poor instruments or too small data sample.

	***TC* OLS**	***C* OLS**	***TC* GMM**	***C* GMM**
λ_1	0.9540***	0.8363***	0.8319*	0.5121
	(0.2554)	(0.2320)	(0.4398)	(0.4293)
λ_2	1.1323***	0.9776***	1.4684***	1.3975***
	(0.2865)	(0.2602)	(0.4158)	(0.3931)
F-test	27.4128	25.1290	17.4526	13.4610
Fp	[0.000]	[0.000]	[0.000]	[0.000]
J	-	-	9.3252	7.7907
Jp	-	-	[0.230]	[0.351]
R-squared	0.6463	0.6262	0.608	0.5635

Table 3 Estimates of equations (3) and (5) in period 1999 – 2010

*Note: μ and ϕ included, not shown. The expanding GMM instruments are lagged 2-3 years. F-test: (H_0: $\lambda_1 = \lambda_2$). The test of restrictions is Sargan's test of overidentifying restrictions (H_0: instruments are valid) with J-statistic and p-value in square brackets with the chi-square distribution. Robust standard errors are in parenthesis, ***, ** and * denote statistical significance at 1%, 5% and 10% respectively.*

4 Conclusion

Using the panel data analysis it can be concluded that PIH-LCD is inconsistent with aggregated annual data from the period 1999 – 2010 for the Baltic States: Estonia, Latvia and Lithuania. We compared the estimates of the share of "rule-of-thumb" consumption λ for the consumption of non-durables and services with the estimates of λ for total consumption. We established that the share of "rule-of-thumb" consumption is very essential in the Baltic States and it is 81% for total consumption and 74% for consumption without the durable component. As the difference between above-mentioned estimates is not very great, then it is possible to use the data of total consumption in the empirical research in the case where the durability data are not available. Additionally it must be admitted that the estimates of λ contain certain ambiguity, because of the very short time period for trustworthy estimation. However, the truthfulness of our estimates is supported by the above-mentioned results of Kukk et al. [14], who also established that the fraction of "rule of thumb" consumption was approximately 75% in the period 2002 – 2007 for Estonia.

Analysis of the asymmetric effects shows that the most probable reason of the rejection of PIH-LCD in the Baltic countries is the "opposite" asymmetry. Both myopia and liquidity constraint are inconsistent with our results. Under myopia, consumption should respond symmetrically to the changes in disposable income, but under liquidity constraints the consumption should be more sensitive to the income increase rather than to decrease. There was a loan boom in these countries in the period 1999 – 2007 before recession. It was very popular among the households to borrow for home mortgages and car leasing. This indicates that the liquidity constraint is not a very probable reason for the rejection of PIH-LCD. The last assertion is also confirmed by the data of

total consumption were the durable component is presented. The consumption for non-durables and services shows that the sign of positive change of disposable income are insignificant and only for the OLS estimation, what is biased, shows the presence of "opposite asymmetry". Shea [21] argues that under myopia non-durables consumption may be also more sensitive to income declines rather to increases. However, the data of total consumption has not confirmed this assertion for the Baltic States

Acknowledgements

This research was supported by the Estonian Science Foundation grant 8627.

References

[1] Baltagi, B.: *Econometric Analysis of Panel Data.* 4th ed., John Wiley & Sons Inc., Chichester, 2008.
[2] Campbell, J., and Mankiw, G.: Consumption, income and interest rates: reinterpreting the time series evidence. *NBER Macroeconomics Annual* **4** (1989), 185–246.
[3] Campbell, J., and Mankiw, G.: Permanent income, current income, and consumption. *Journal of Business & Economic Statistics* **8** (1990), 265–279.
[4] Campbell, J., and Mankiw, G.: The response of consumption to income. A cross-country investigation. *European Economic Review* **35** (1991), 723–767.
[5] Chyi, Y., and Huang, C.: An empirical study of the "rule-of-thumb" consumption model in five East Asian countries. *Applied Economics* **29** (1997), 1271–1282.
[6] Diamond, P.: National debt in a neoclassical growth model. *American Economic Review* **55** (1965), 1126–1150.
[7] Filer, L., and Fisher, J.: Do liquidity constraints generate excess sensitivity in consumption? New evidence from a sample of post-bankruptcy households. *Journal of Macroeconomics* **29** (2007), 790–805.
[8] Friedman, M.: *A theory of the consumption function*. Princeton University Press, Princeton and Oxford, 1957.
[9] Fuhrer, J.: Habit formation in consumption and its implications for monetary-policy models. *American Economic Review* **90** (2000), 367–390.
[10] Gomes, F., and Paz, L.: Consumption in South America: myopia or liquidity constraints? *Economia Aplicada* **14** (2010), 129–145.
[11] Hall, R.: Stochastic implications of the Life Cycle - Permanent Income Hypothesis: theory and evidence. *Journal of Political Economy* **86** (1978), 971–87.
[12] Issler, J., and Rocha, F.: Consumo, restrição a liquidez ebem-estar no brasil. *Economia Aplicada* 4 (2000), 637–665.
[13] Kiley, M.: Habit persistence, non-separability between consumption and leisure, or rule-of-thumb consumers: which accounts for the predictability of consumption growth? *The Review of Economics and Statistics* **92** (2010), 679–683.
[14] Kukk, M., Kulikov, D., and Staehr, K.: *Consumption sensitivities in Estonia: income shocks of different persistence.* Estonian Bank Working Paper Series nr. 3, 2012.
[15] Levin, A., Lin, C.F., and Chu, C.-S. J.: Unit root tests in panel data: asymptotic and finite sample properties. *Journal of Econometrics* **108** (2002), 1–24.
[16] Reis, R., Issler, J., Blanco, F., and Carvalho, L.: Renda permanente e proupança precaucional: evidências empíricas para o brasil no passado recente. *Pesquisa e Planemajento Econômico* **28** (1998), 233–272.
[17] Ridala, S.: "Rule-of-thumb" consumption in the Baltic States. *Journal of International Scientific Publications: Economy & Business* **6** (2012), forthcoming.
[18] Sahm, C., Shapiro, M., and Slemrod, J.: *Check in the mail or more in the paycheck: does the effectiveness of fiscal stimulus depend on how it is delivered?* NBER Working Paper Series nr. 16246, 2010.
[19] Sarantis, N., and Stewart, C.: Liquidity constraints, precautionary saving and aggregate consumption: an international comparison. *Economic Modeling* **20** (2003), 1151–1173.
[20] Shea, J.: Union contracts and the Life-Cycle/Permanent-Income Hypothesis. *American Economic Review* **85** (1995), 186–200.
[21] Shea, J.: Myopia, liquidity constraints, and aggregate consumption. *Journal of Money, Credit and Banking* **27** (1995), 798–805.
[22] Souleles, N.: The response of household consumption to income tax refunds. *American Economic Review* **89** (1999), 947–58.
[23] Verbeek, M.: *A Guide to Modern Econometrics*. John Wiley & Sons Inc., Chichester, 2012.
[24] Weber, C.: "Rule-of-thumb" consumption, intertemporal substitution, and risk aversion. *Journal of Business and Economic Statistics* **18** (2000), 497–502.

Possibilities of control congested intersections controlled by traffic lights

Michal Rusek[1]

Abstract. With the increasing values of vehicles are becoming increasingly frequent lack of infrastructure capacity that is used by vehicles. This phenomenon is most evident in urban transport infrastructure, where vehicles concentrations are highest. Typical examples are light-controlled intersections, which are often heavily loaded, which is reflected in the congested area in front of intersections. Solving of this problem is very difficult but very necessary, because the elimination of this undesirable phenomenon in traffic has resulted in reducing of economic losses by congestion, more efficient using of transport and last but not least improving the quality of life in the city. In the past has been shown that using linear programming methods can be designed signal plans, which can effectively control the intersections. But it was at intersections which are not congested. This paper deals the possibilities of designing of signal plans using linear programming in a situation where the intersections are congested. Congestion is reflected by the fact that is not ensured that all vehicles that came to the intersection in one cycle, it leaves during the same cycle. In the beginning of this article are described the possible ways of solving depending on the requirements of traffic flows, followed by a description of mathematical models which can solve this problem. The theoretical description is supported by experiments carried out and in conclusion are summarized the advantages and disadvantages of the approach.

Keywords: intersection signal plan, traffic flows, congestions, linear programing.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

In real road traffic are very often problems, that are caused by insufficient capacity of transport facilities that were built in the past and there wasn't assumption of such high growth in the number of vehicles.

The most significant problems are encountered in the cities. Very common phenomenon, to which it is possible to meet, is creation of the congestions. This problem is caused by the fact that entry into the intersection during period runs an average of more vehicles than it is able to release the intersection

The presented paper discusses how to secure control of intersections which are congested and are unable to ensure that all vehicles come to the intersection during one cycle, it has been able during the same cycle also leave.

Management will be implemented through the intersection signal plans. Signal is a plan document, stating at which point will have different traffic flows in the intersection green signal. Creating of signal plan will be carried out using linear programming.

Mathematical models for creating signal plans presented in the past always assumed that the vehicles which come to the intersection, during the same cycle can leaves [1], [2], [4]. In congested intersections where congestion is created, this assumption can not be secure and is not therefore possible to use these mathematical models. In the present paper will present a mathematical model with which it is possible to generate a signal plans for congested intersections. For this mathematical model will also present the possible modifications to improving monitored parameters.

2 Mathematical model for designing intersection signal plans

Let the traffic flows entering the intersection constitute a set I. Each traffic flow i from the set I is specified by the intensity f_i. During the red signal vehicles in the flow create the queue and after the beginning of the green signal vehicles leave the intersection with saturated intensity f_i^s. Technical standards set value τ_i for each type of traffic flow and this value is the minimum time for green signal of the flow i. [1]

[1] VSB-Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17.listopadu 15, 70833 Ostrava-Poruba, Czech Republic, michal.rusek@atlas.cz

For every two collision flows i and j we assume the minimum time of delay between the end of the green signal of flow i and the beginning of the green signal of flow j as m_{ij}. The period of intersection signal plan will be considered as the interval $<0, t_{max}>$. We introduce variables z_i to model missing length of green signal, necessary for the release of all vehicles of the flow i and variables x_i and y_i to model the start and end of the green signal of the flow i. We also introduce phases of the set $K = \{F_1, F_2, ..., F_r\}$, where F_k is the set of non-collision traffic flows, i.e. Traffic flows which vehicles can pass through the intersection simultaneously [1].

Now can be presented basic mathematical model. Model is based on models presented for example in [4], [5]. This model can be written in the form:

$$Minimize \sum_{i \in I} f_i^s \cdot z_i \tag{1}$$

$$Subject\ to\ \ y_i - x_i + z_i \geq \left(\frac{f_i t^{\max}}{f_i^s} \right) \quad for\ \ i \in I \tag{2}$$

$$y_i - x_i \geq \tau_i \quad for\ \ i \in I \tag{3}$$

$$x_j - y_i \geq m_{ij} \quad for\ k = 1,..., r-1, s = k+1..r,\ i \in F_k,\ j \in F_s,\ i, j - collision \tag{4}$$

$$x_j - y_i \geq m_{ij} - t_{\max} \quad for\ k = 2,..,r,\ s = 1,..,k-1,\ i \in F_k,\ j \in F_s,\ i, j - collision \tag{5}$$

$$y_i \leq t_{\max} \quad for\ i \in I \tag{6}$$

$$x_i \leq t_{\max} \quad for\ i \in I \tag{7}$$

$$x_i \in Z^+ \quad for\ i \in I \tag{8}$$

$$y_i \in Z^+ \quad for\ i \in I \tag{9}$$

$$x_i \leq y_i \quad for\ i \in I \tag{10}$$

In this model, expression (1) represents the objective function, in which is performed minimizing of the number of vehicles which was not released by intersection in one cycle and remain until the next cycle. The expression (2) is the condition which ensures that the time of green signal has been so long how is needed to release the most vehicles how is possible pass through intersection. The variable z_i in this condition is presented because if it was not possible to release all vehicles. Thanks this fact, without this variable would not satisfy the condition of equality then there wasn't solution and therefore variable z_i modeling missing time off signal ensures that the condition is always satisfied. The expression (3) is a condition which ensures that the length of the green signal takes at least reaching a value that is set by standard for the type of traffic flow. Expressions (4) and (5) ensure that required split times between collision traffic flows will be observed. Condition (6) ensures that value of the end of green signals don't exceed the length of the cycle. Similarly, condition (7) ensures that starts of green signals do not exceed the length of the cycle. The expression (10) ensures that beginning of green signal preceded the end of green signal and there was no violation of the natural order of values. Expressions (8) and (9) represent the obligatory conditions of the model - Domains of variables.

In some cases, however, can be required to minimize the number of vehicles that remain before the intersection, conducted primarily for certain preferred traffic flows (e.g. for traffic flows on the main direction or on the busiest traffic flows). In this case, it is possible to edit the original objective function (1) and replace it with the modified objective function (11). Also condition (2) is needed to replace by the condition (12).

$$Minimize \sum_{i \in I^P} f_i^s \cdot z_i \tag{11}$$

$$y_i - x_i + z_i \geq \left(\frac{f_i t^{\max}}{f_i^s} \right) \quad for\ \ i \in I^P \tag{12}$$

The expression (11) represents the objective function, in which is minimize the number of vehicles which wasn't released and remain until the next cycle. The difference with the objective function (1) of basic model is just that the sum isn't over all traffic flows, but only over a preferred traffic flows from the set of preferred traffic flows I^P. The set of all traffic flows I is composed of subsets of preferred I^P traffic flows and traffic flows of non-preferred set I^N, thus $I = I^P + I^N$. The expression (12) is a condition which ensures that the time of green signal has been so long how is needed to release the most vehicles how is possible pass through intersection, but now this condition is applied only at preferred traffic flows from the set I^P.

Now is also possible modification of objective functions of both models. For many tasks as optimization criterion is used total waiting times in intersection per cycle. This optimization criterion is minimized. While it is done in the objective function to minimize the number of vehicles remaining in the intersection to the next cycle, it is also possible to carry out and minimize the total waiting times in way that the original objective function is complement by objective function, in which is optimizing total waiting time in the intersection.
The original objective function (1), for the first model now has the form (13).

$$Minimize \sum_{i \in I} f_i^s \cdot z_i + 0.00001 \cdot \left(\sum_{i \in I} 0.5 * \left(\frac{f_i * f_i^s}{f_i^s - f_i} \right) * u_i^2 \right) \tag{13}$$

And similarly, the original objective function (11), for the second model with the thinking of preferred traffic flows is now the preferred in form (14).

$$Minimize \sum_{i \in I^P} f_i^s \cdot z_i + 0.00001 \cdot \left(\sum_{i \in I} 0.5 * \left(\frac{f_i * f_i^s}{f_i^s - f_i} \right) * u_i^2 \right) \tag{14}$$

Is also needed to add to the list of variables and constants entering the model variable u_i, modeling the length of the red signal for the i-th traffic flow. This variable is then replaced by piecewise linear function [2]. Also, it is necessary to add to the objective function also any sufficiently small constant, to show which optimization criterion is significant.

3 Solving method and numerical experiments

It is now possible to write presented models to the form which is required by optimization software [3] and subsequently solved. The following numerical experiments will now be performed in order to verify that the presented mathematical models are applicable to solve practical problems of various ranges. As a test file will be used group of 17 intersections located in the city of Ostrava.

Due to the fact that all the intersections that were available are in category of non congested intersections, an adjustment was made. The initial intensity was increased by 10, 20 and 30%. Each intersection was thus solved for the three values of intensities.

Mathematical models that have been addressed are treated with compound objective functions (13) and (14). They were both designed versions, thus minimizing the number of vehicles remaining to next cycle in a situation where all traffic flows have the same priority (Model 1) and in a situation where some traffic flows are preferred (Model 2).

The results are shown in Table 1 Presented are the results for only some intersections because of extensive output data. The table contents for each intersection numbers of vehicles at different intensities and variations of models remain in the intersection to the next cycle. Due to the fact that the numbers of vehicles are compared with a preference for certain tasks and traffic flows without preference in the table are values only of queues of preferred traffic flows to see that the model with preferences can reduce the numbers of vehicles in their queues.

Intersection 2070		
Value of intensity:	**Model 1**	**Model 2**
	Number of vehicles in monitored traffic flows remaining to next cycle	Number of vehicles in monitored traffic flows remaining to next cycle
Intensity +10%	P1:0; P4:0	P1:0; P4:0
Intensity +20%	P1:0; P4:0.422	P1:0; P4:0
Intensity +30%	P1:0; P4:0.4155	P1:0; P4:0
Intersection 1006		
Value of intensity:	**Model 1**	**Model 2**
	Number of vehicles in monitored traffic flows remaining to next cycle	Number of vehicles in monitored traffic flows remaining to next cycle
Intensity +10%	P2:0; P4:0 P6:0; P7:0	P2:0; P4:0 P6:0; P7:0
Intensity +20%	P2:0.8883; P4:0 P6:0; P7:0	P2:0; P4:0 P6:0; P7:0
Intensity +30%	P2:2.33733; P4:0 P6:0; P7:0	P2:0; P4:0 P6:0; P7:0
Intersection 1018		
Value of intensity:	**Model 1**	**Model 2**
	Number of vehicles in monitored traffic flows remaining to next cycle	Number of vehicles in monitored traffic flows remaining to next cycle
Intensity +10%	P1:1.068; P2:0	P1:0; P2:0
Intensity +20%	P1:2.256; P2:0.296	P1:0; P2:0
Intensity +30%	P1:2.444; P2:0.404	P1:0; P2:0
Intersection 3007		
Value of intensity:	**Model 1**	**Model 2**
	Number of vehicles in monitored traffic flows remaining to next cycle	Number of vehicles in monitored traffic flows remaining to next cycle
Intensity +10%	P3:0; P6:0.27016	P3:0; P6:0
Intensity +20%	P3:0; P6:1.79472	P3:0; P6:0
Intensity +30%	P3:0; P6:5.31928	P3:0; P6:0
Intersection 4006		
Value of intensity:	**Model 1**	**Model 2**
	Number of vehicles in monitored traffic flows remaining to next cycle	Number of vehicles in monitored traffic flows remaining to next cycle
Intensity +10%	P2:0.3853; P5:0.1769	P2:0; P5:0
Intensity +20%	P2:2.1476; P5:1.3748	P2:0; P5:0
Intensity +30%	P2:2.9099; P5:4.0727	P2:0; P5:0

Table 1 Table of results

4 Conclusion

Results in the Table 1 shows that with thinking of the preference of certain traffic flows is possible, by a model in which the numbers of vehicles are priority optimize in preferred flows, to eliminate number of vehicles that remain in the intersection to the next cycle.

But there is no guarantee that the elimination may be possible always. The mathematical model can't eliminate the number of vehicles remaining to the next cycle always, because if there isn't already available free capacity, or for non-preferred traffic flows can't be reduce green signal and extend it at preferred traffic flow (thus reducing the number of vehicles waiting in the next cycle) the number of vehicles can't be eliminated. Although the results in Table don't show this situation and always managed to eliminate the number of vehicles, carrying out experiments in a few cases wasn't possible to reach elimination.

Conversely, if the elimination of remain vehicles before the intersection is successful, it is necessary to realize that this is the expense of other traffic flows, and therefore these are contrary the number of vehicles remaining to the next cycle are growing.

The present paper shows that it is possible thanks to linear programming solve congested intersections and after making of the necessary adjustments can be priority made an optimization of preferred traffic flows and

apart from the optimization of number of vehicles remaining in the next cycle is also possible to perform optimization of total waiting times in intersection.

Acknowledgements

This article was supported by the grant project of the Faculty of Mechanical Engineering Technical University of Ostrava, No. SP2012/113 "Development of new methods to support planning and management of transportation processes".

References

[1] Černý, J. and Kluvánek P.: *Základy matematickej teórie dopravy*, VEDA, Bratislava, 1991.
[2] Koháni, M. and Rusek M. Comparison of Different Approaches to the Control of Cyclic processes. In: *book of papers Mathematical Methods in Economics 2011*. Professional Publishing, Prague, 2011.
[3] XPRESS-Mosel *User guide*. Dash Associates, Blisworth, 2005.
[4] Janáček, J. and Koháni, M.: Waiting Times in Public Transport and Their Modelling for IP-Solver. *Communications–Scientific Letters of the University of Žilina*, Vol. 12,No. 3A, Žilina, 2010
[5] Koháni, M. and Rusek M.: The impact of "red point" for vehicle waiting times in intersections. In: *Journal of Information, Control and Management Systems*, Vol. 8, No. 3, Žilina, 2010.

Quantification of framing effect using ANP

Jan Rydval[1]

Abstract. When it comes to decision making the most influential components are the quality of the information and the size of the framing effect. In the decision-making process the framing effect arises as a set of opinions and expectations of involved subjects. Everyone making a decision has his own preferences and expectations that create his own unique view (frame). These unique views or frames may negatively influence the information sharing in the decision-making processes and therefore the analysis of the framing effect is important.
This paper shows how to use Analytic Network Process to do the determination and quantification of the framing effect and it shows the advantages of using this method. The result of this approach will be the set of weights of all viewpoints included into the frames. Values of these weights can serve to analyse the framing effect, because all subjects in the decision-making process will receive the information about the importance of different viewpoints and then they can rearrange their opinions and expectations. It will have positive impact on the overall outcome of the decision-making process too.

Keywords: Framing Effect, Analytic Network Process, Pairwise Comparison, Product Packaging

JEL Classification: C44
AMS Classification: 90C29

1 Introduction

The integral part of any decision-making processes is the information receiving process. As Fagley, Coleman and Simon [3] mention decision-making is influenced by the quality of the information and by the effect of information distortion (framing effect). And as mentioned by Tversky and Kahneman [15] the framing effect included in some information can significantly influence decisions. There are various views on a particular issue in decision process. This various views (or frames) may create several issues. As Bishop [1] believes that if information is not sorted according to its relevance because we cannot properly decide who has the most important view, we may face the problem of being overloaded with too much information resulting in either poor information acquisition or the whole process is very time consuming and thus very ineffective. On the contrary to this situation preferring the certain point of view we may lose the information needed for successful decision making. To limit these negative frames we need firstly to define and understand them, secondly, as pointed out by Druckman [2], we need to evaluate them, and thirdly we need to use the appropriate method to reduce them. For this kind of evaluation as Fagley, Coleman and Simon [3] write we need to know the importance of various frames and included points of view. These frames can be analyzed using multiple attribute decision-making methods.

The typical example of the frame's impact on decision is the information written on the product package. As Kozel [4] analyses in more details, the packaging inevitably influences our purchase decisions. In the current theories the packaging is presented as one of the forms of the marketing communication. According to Lindsey-Mullikin and Petty [5], the product packaging can attract our attention; affect our emotions, but on the other hand, by its information value it can contribute to the rational purchase decision. The main reason behind this is the fact, that consumers expect the products to meet their expectations and prefer them according to their own preferences. Each consumer has his unique view when buying a specific product (his unique way of perceiving the situation) based on his personality. His purchase decision is influenced by the preferences and expectations; we are talking about the framing effect of the decision situation, which discusses in detail Rydval [8], [9].

The Analytic Network Process is one of the multiple criteria decision making methods. It decomposes decision problems into a network of smaller parts (sub-problems) that can more easily by analysed and evaluated. It is specific for this method that the human judgment is involved. (Saaty [11])

1 Czech University of Life Sciences Prague, Faculty of Economics and Management, Department of Systems Engineering, Kamýcká 129, 165 21 Praha 6 – Suchdol, rydval@pef.czu.cz.

The goal of this paper is to quantify the framing effect, its main set of viewpoints and their components, and their evaluation using the Analytic Network Process and to show the impact of network on the evaluation of framing effect, opposite to a hierarchy.

2 Materials and methods

2.1 Framing effect

Individual decisions are influenced by the presented information and by the formulation of problems. Each subject has his own preferences and expectations that create his own view (frame). The framing effect is made up of these frames. Tversky and Kahneman [15] pointed out the framing effect influences the way of information interpretation or misinterpretation, so it may influence decision-making significantly. As Rydval and Hornická [7] mentioned we can therefore define the framing effect as a set of preferences and expectations of involved subjects belonging to a particular decision-making problem. To quantify framing effect, as indicated Rydval [8] , [9], the methods for quantifying preferences of decision maker can be used. Quantification of the framing effect in education is discussed by Rydval and Brožová [6].

2.2 Saaty's method of pairwise comparison

It is a quantitative pairwise comparison method for the criteria, which analyzes in detail Saaty [12, 13]. A nine point scale is provided to quantify pairwise importance of criteria and it is possible to use intermediate values (values 2, 4, 6, 8):

Intensity of Importance	Definition	Explanation
1	Equal Importance	Two activities contribute equally to the objective
2	Weak or slight	
3	Moderate importance	Experience and judgement slightly favour one activity over another
4	Moderate plus	
5	Strong importance	Experience and judgement strongly favour one activity over another
6	Strong plus	
7	Very strong or demonstrated importance	An activity is favoured very strongly over another; its dominance demonstrated in practice
8	Very, very strong	
9	Extreme importance	The evidence favouring one activity over another is of the highest possible order of affirmation

Table 1 The fundamental scale of absolute numbers

Expert compares each pair of criteria and he records the size of the preferences of the *i* criterion to the *j* criterion in the Saaty matrix $S = (s_{ij})$:

$$S = \begin{pmatrix} 1 & s_{12} & \cdots & s_{1n} \\ 1/s_{12} & 1 & \cdots & s_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ 1/s_{1k} & 1/s_{12} & \cdots & 1 \end{pmatrix} \qquad (1)$$

If the *i* and *j* criterion are equal then $s_{ij} = 1$; If the *i* is slightly stronger than *j* criterion then $s_{ij} = 3$; If the *i* is stronger than *j* criterion then $s_{ij} = 5$; If the *i* is much stronger than *j* criterion then $s_{ij} = 7$; If the *i* is absolutely stronger than *j* criterion then $s_{ij} = 9$; If *j* is preferred instead of *i*, inverted values are entered into the Saaty matrix ($s_{ij} = 1/3$ for a weak preference, $s_{ij} = 1/5$ for a strong preference, etc.).

It is a square matrix of order $n \times n$, reciprocal, i.e. $s_{ij} = 1/s_{ji}$, and expresses actually the estimate of weights of i and j criterion. On the diagonal of Saaty matrix values are always one (each criterion is equivalent to itself).

The elements of this matrix are not usually perfectly consistent, $s_{hj} = s_{hi} \times s_{ij}$ is not valid for all $h, i, j = 1, 2, ..., n$. If we compiled a matrix $V = (v_{ij})$, whose elements would be real weights ($v_{ij} = v_i / v_j$), for elements of this matrix, the above condition is applied. The rate of consistency is measured by a consistency index defined by Saaty:

$$I_S = \frac{l_{\max} - n}{n-1}, \qquad (2)$$

where l_{max} is the largest eigenvalue of Saaty's matrix and n is the number of criteria. Saaty's matrix is considered to be sufficiently consistent if $I_S < 0.1$.

To determine the weights Saaty suggested several wals we can use to estimate the weights vj. The most commonly used procedure to calculate the weights as normalized geometric mean of the rows of the Saaty's matrix (logarithmic least squares method). Calculate the value bi as the geometric mean of the rows of the Saaty's matrix:

$$b_i = \sqrt[n]{\prod_{j=1}^{n} s_{ij}} \qquad (3)$$

The weights are calculated by normalization of values b_i:

$$v_i = \frac{b_i}{\sum_{i=1}^{n} b_i} \qquad (4)$$

Saaty's method can be used not only to determine the preferences between the criteria, but also among the variants.

2.3 Analytic network process

The Analytic Network Process (ANP) is a generalization of the Analytic Hierarchy Process (AHP), by considering the dependence between the elements of the hierarchy. Many decision problems cannot be structured hierarchically because they involve the interaction and dependence of higher-level elements in a hierarchy on lower level elements. Therefore, ANP is represented by a network, rather than a hierarchy. (Saaty [10], [12]), [13]))

The basic elements of the ANP method are following:

- The first step of ANP is based on the creation of a control network which describes dependency among decision elements. The ANP allows
  - inner dependence within a set (clusters) of elements, and
  - outer dependence among different sets (clusters).
- In the second step pairwise comparisons of the elements within the clusters and among the clusters are performed according to their influence on each element in another cluster or elements in their own cluster. So the ANP prioritizes not only decision elements but also their groups or clusters as it is often the case in the real world. The consistency of these comparisons has to be controlled.
- The third step consists of the supermatrix construction. The priorities derived from the pairwise comparisons are entered into the appropriate position in this supermatrix. This supermatrix has to be normalized using clusters weights.
- In the fourth step the limiting supermatrix is computed and global preferences of decision elements are obtained. These preferences serve as the best decision selection or for the purpose of analysis of preferences of decision-making elements. (Saaty [10], [11])

2.4 SuperDecisions software

This method is carried out by the SuperDecisions software system (SuperDecisions [14]). The SuperDecisions software implements the ANP developed by Dr. Thomas Saaty. The program was written by the Creative Decisions Foundation.

3 Results and discussion

3.1 A case study – Measuring the framing effect using the ANP through the analysis of information on pork

Information published on the product packaging is very important when analyzing consumer behavior and it has a major impact on consumer demand. Consumer's own criteria create a consumer's view on the product purchase. Because it participates in the formation of preferences, it is necessary to adequately define and quantify it. Criteria, on which customer put emphasis when buying products, can be defined using questionnaires. However, to obtain information appropriate for the quantification of consumer's criteria, we must properly structure questionnaires.

This case study follows the case study analyzing the product information using AHP, in which Rydval [9] analyzes information on the packaging of pork. The survey was carried out in Prague and the responses were obtained from 86 customers and 9 meat sellers. Average pork meat contains from 35 to 55% fat, while lean pork only from 15 to 20% (Caloric tables). Purchasing pork can be simply characterized as multi-criteria analysis model of variants. The consumer decides between variants based on his criteria. He decides which one he will buy. On the other hand, the producer bases his decision about the information he publishes about the product to attract the customer on the existing legislation and his own criteria. It is important for the producer to know not only the legislation and his preferences when publishing information, but also the consumer's preferences. These two sets of preferences, respectively two views on the same issue, must be defined in an appropriate way and based on the way we want to present our product. For a situation where the consumer decides between multiple variants the crucial are the criteria, on which customer bases his decisions. The criteria can be determined by questionnaire. We can use the same approach for producer.

In this case study, the important for consumers when purchasing meat, were the following criteria:

- Price (the price of the product in CZK)
- Quality (measured by the meat texture according to standards ISO 11036, 1997 and by weight loss during heat treatment in %)
- Manufacturer (the name of manufacturer or importer respectively supplier’s name and the country of origin)
- Availability (measured by the availability of the product in retail network)
- Production Ecology (the product is labeled as "organic product" with organic logo, measured by the criteria specified in Act No. 242/2000 on organic agriculture, as amended)

When considering what information about the product to publish, the following criteria were the most important for the producers (sellers):

- Quality (measured by the meat texture according to standards ISO 11036, 1997 and by weight loss during heat treatment in %)
- The originality of the product (a combination of technological processing of the product and the creative adaptation of the product packaging)
- Price (the price of the product in CZK)
- Manufacturer (the name of manufacturer or importer respectively supplier’s name and the country of origin)
- Production Ecology (the product is labeled as "organic product" with organic logo, measured by the criteria specified in Act No. 242/2000 on organic agriculture, as amended)

Rydval [9] used for the criteria quantification the AHP model in his case study. Figure 1 shows the structure. In this case, the ANP model is used to quantify the criteria. The structure of this model is shown in Figure 2

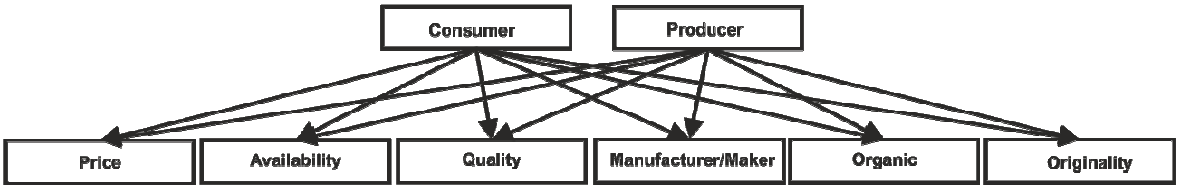


Figure 1 Hierarchical structure of decision-making model; Information on pork

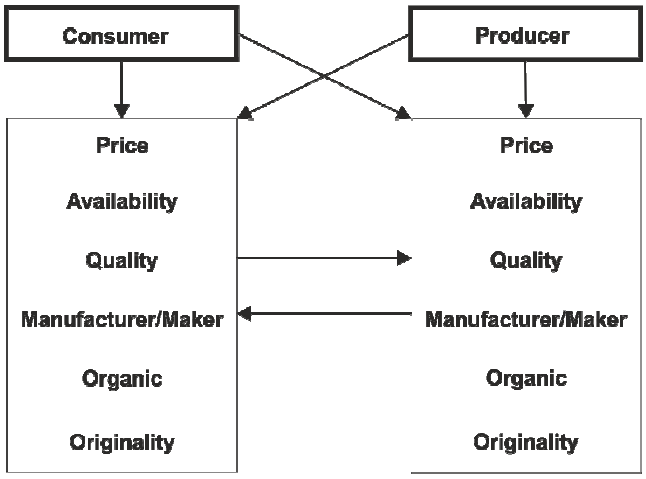


Figure 2 Network structure of decision-making model; Information on pork

Based on values obtained from the questionnaires and using Saaty methods of pairwise comparisons of criteria, weights of particular customers' and producers' criteria were compiled. In a case study Rydval [9] conducted pairwise comparisons based on the hierarchical structure of the problem, as shown in Table 2 in the column "AHP Structure". In this case study pairwise comparisons were conducted based on the network structure of the problem, as shown in Table 2 in the column "ANP Structure".

	AHP Structure		**ANP Structure**	
	Frame of Consumer	**Frame of Producer**	**Frame of Consumer**	**Frame of Producer**
Price	0,51	0,18	0,65	0,38
Quality	0,24	0,43	0,28	0,36
Originality	0,03	0,21	0,02	0,14
Producer	0,13	0,10	0,09	0,05
Organic	0,04	0,06	0,01	0,03
Availability	0,05	0,02	0,01	0,04

Table 2 Criteria frames using AHP and ANP

Criteria frames (each subject of the particular issue has its own quantified criteria frame) play an important role when choosing product to purchase. Therefore, as the seller we decide to present not only by us preferred quality, but mainly the price the consumer prefers. But we still can see that quality of the purchased product is very important for the consumer.

This case study shows how ANP helps to model the user's preferences, and serves to quantify the criteria for his purchase. Thus it helps to decide on what information to show on the product packaging and make the product more attractive for the consumers. ANP may thus serve as a supporting tool in the marketing strategy. As the table shows the results in the AHP and the ANP structure are very different. This is mainly due diametrical difference in the structure of models. Individual elements of AHP structure are independent of each other at various levels, while elements in the ANP structure influence each other. Another big advantage of the ANP model is the possibility of obtaining weights of individual criteria from the perspective of other criteria, not only from the perspective of the assessor. Therefore, in more complex situations ANP models serve better than AHP models.

We can therefore define the framing as a set of weights of individual aspects (preferences and expectations) affecting the decision of the decision maker that can be quantified using multicriteria analysis of variants. In this case, because of the complexity of the situation, the method ANP method seems to be a suitable.

4 Conclusion

This paper deals with the factors affecting our rational thinking, with our ability to make rational decisions, and in particular it serves to explain the framing effect in decision-making process and its quantification using the ANP method. The framing effect affects the ability to reach the rational choice mostly in a negative way and therefore it can make decision-making processes very difficult and minimize the quality of decisions. It may have fatal consequences and it can negatively affect passing the information when making a purchase decision.

In its case study this paper shows the rise of the two dominant frames that influence the outcome of the purchase decision in a market place. Using the ANP method, views (frames) from the whole perspective were created. In these new frames all previous dominant frames are aggregated and their influence is minimized. It can help the decision maker to move closer to a more suitable rational decision and positively influence the overall outcome of the purchasing process.

ANP is a more general form of the AHP used in multi-criteria decision analysis. AHP structures a decision problem into a hierarchy with a goal, decision criteria, and alternatives, while the ANP structures a decision problem as a network. Both of them use a system of pairwise comparisons to measure the weights of the components of the structure, and finally to rank the alternatives in the purchase decision. In the AHP method, each element in the hierarchy is considered to be independent of all the others. The decision criteria are considered to be independent of each other, and the alternatives are considered to be independent of the decision criteria and of one another. But in many real-world cases, there is interdependence among the items and the alternatives. ANP does not require independence among elements, so it can be used as an effective tool in these cases. Another big advantage of the ANP model is the possibility of obtaining weights of individual criteria from the perspective of other criteria, not only from the perspective of the assessor. Therefore, in more complex situations ANP models serve better than AHP models. Therefore the weights determined by AHP for the particular preferences vary from the ones determined by ANP. This can be clearly seen in the case of the preferences "Price" and "Quality". And because of the interdependence among the items and the alternatives, where there are preferences and expectations dependent on each other, it is better to use ANP in these situations.

The more detailed interpretation of its results can be achieved when using the ANP method. The results can be interpreted from the view of individual preferences of participating subjects. The purchasing process can be adjusted according to the results of this case study.

Acknowledgements

The research is supported by the Internal Grant Agency of the University of Life Sciences Prague – project IGA PEF 20121032 and project IGA PEF 201111160036.

References

[1] Bishop, J.: *Chapter 15 Drinking from the Fountain of Knowledge: Student Incentive to Study and Learn – Externalities, Information Problems and Peer Pressure*. Handbook of the Economics of Education, 2006, Volume 2, Pages 909-944, ISBN: 978-0-444-52819-3

[2] Druckman, J. N.: *Evaluating framing effects.* Journal of Economic Psychology, 2001, Volume 22, Issue 1, ISSN 0167-4870.

[3] Fagley, N. S., Coleman, J. G., Simon A. F.: *Effects of framing, perspective taking, and perspective (affective focus) on choice.* Personality and Individual Differences, 2010, Vol: 48, Issue: 3, ISSN 0191-8869.

[4] Kozel, R.: *Moderní marketingový výzkum*. Grada, Praha, 2006.

[5] Lindsey-Mullikin, J., Petty, R. S.: *Marketing tactics discouraging price search: Deception and competition.* Journal of Business Research, 2011, Volume 64, Issue 1, ISSN 0148-2963.

[6] Rydval, J., Brožová, H.: Quantification of Framing effect in education Process using ANP. Proceedings of Efficiency and Responsibility in Education International Conference 2011, Prague, CULS, 2011, ISBN 978-80-213-2183-0.

[7] Rydval, J., Hornická A.: *Reduction of framing effect in transport logistic using decision support system.* Sborník - Think Together 2011, Praha, ČZU, 2011, ISBN 978-80-213-2169-4

[8] Rydval, J.: *Reducing The Framing Effect in Decision Processes*. Proceedings of UCOLIS 2010 - University Conference in Life Sciences – Proceedings, Praha, ČZU, 2010, ISBN 978-80-213-2141-0.

[9] Rydval, J.: *Using AHP for analysing product information*. Proceedings of Think Together 2012, Praha, ČZU, 2012.
[10] Saaty, T. L.: *Decision Making with Dependence and Feedback: The Analytic Network Process*. 1996, ISBN 0-9620317-9-8, RWS
[11] Saaty, T. L.: *Decision Making with Dependence and Feedback: The Analytic Network Process, The Analytic Hierarchy Process Series*. Pittsburgh: 2001, IX, RWS Publications.
[12] Saaty, T. L.: *How to make a decision: The analytic hierarchy process*. European Journal of Operational Research, 1990, Volume 48, Issue 1,
[13] Saaty, T.L.: *The Analytic Hierarchy Process (AHP) for Decision Making and the Analytic Network Process (ANP) for Decision Making with Dependence and Feedback, Creative Decisions Foundation*. 2003
[14] SuperDecisions Software for Decision-Making. URL <http://www.superdecisions.com/>
[15] Tversky, A., Kahneman, D.: *The framing of decisions and the psychology of choice*. Science 211, 1981

Measuring the efficiency in the Czech banking industry: Data Envelopment Analysis and Malmquist index

Iveta Řepková[1]

Abstract. This paper estimates the technical efficiency and the efficiency change in the Czech commercial banks during the period 2001–2010. We applied the Data Envelopment Analysis and the Malmquist index on the data of the Czech banks. We simultaneously use two alternative specifications of DEA approach, specifically CCR model and BCC model that differ in returns to scale assumption. The results of DEA models show that the average efficiency computed under the assumption of constant returns to scale ranges from 49 to 85 % and the average efficiency estimated under the assumption of variable returns to scale ranges from 71 to 87 %. The results of the Malmquist index reached an annual average negative growth of -4.7%. This negative change can be dichotomized into efficiency change and technological change. Technological change reached an average annual negative growth of -5.6% and technical efficiency change improved average by 1%.

Keywords: efficiency, Data Envelopment Analysis, Malmquist index, Czech banking sector.

JEL Classification: G21, C58
AMS Classification: 62P20

1 Introduction

The two general approaches used to assess efficiency of an entity, parametric and non-parametric methods, employ different techniques to envelop a data set with different assumptions for random noise and for the structure of the production technology. The nonparametric methods are Data Envelopment Analysis and Free Disposal Hull, which are based on linear programming tools. The efficiency frontier in nonparametric estimations is formed as a piecewise linear combination of best-practice observations. The main drawback of nonparametric methods is that they are not robust to measurement errors and luck observed in the data. The parametric methods most widely used in empirical estimations are Stochastic Frontier Approach, Distribution Free Approach and Thick Frontier Approach, which assume specific functional form for the cost function or production technology.

The aim of this paper is to estimate the technical efficiency and the efficiency change in the Czech commercial banks during the period 2001–2010. For the estimation we applied the Data Envelopment Analysis (DEA) and Malmquist index (MI) on the data of the Czech banks. The MI is determined in order to investigate the levels of and the changes in the efficiency of the Czech banks over the analyzed period. The DEA measures the relative efficiency of a homogeneous set of decision-making units (DMUs) in their use of multiple inputs to produce multiple outputs. We simultaneously use two alternative specifications of DEA approach, specifically CCR model and BCC model that differ in returns to scale assumption. The structure of the paper is follow. Next section presents methodology and data, the Data Envelopment Analysis and the Malmquist index and selection of variables are described. Section 3 reveals and discusses the estimated results and Section 4 concludes the paper with summary of key findings.

2 Methodology and data

2.1 Data Envelopment Analysis

The Data Envelopment Analysis is a mathematical programming technique that measures the efficiency of a decision-making unit relative to other similar DMUs with the simple restriction that all DMUs lie on or below the efficiency frontier [9]. DEA also identifies, for inefficient DMUs, the sources and level of inefficiency for each of the inputs and outputs [4].

The CCR [3] model presupposes that there is no significant relationship between the scale of operations and efficiency by assuming constant returns to scale (CRS) and it delivers the overall technical efficiency. The CRS assumption is only justifiable when all DMUs are operating at an optimal scale. However, firms or DMUs in

[1] Silesian University in Opava, School of Business Administration in Karviná, Department of Finance, Univerzitní náměstí 1934/3, Karviná 733 40, repkova@opf.slu.cz.

practice might face either economies or diseconomies to scale. Thus, if one makes the CRS assumption when not all DMUs are operating at the optimal scale, the computed measures of technical efficiency will be contaminated with scale efficiencies. [1] extended the CCR model by relaxing the CRS assumption. The resulting BCC [1] model was used to assess the efficiency of DMUs characterized by variable returns to scale (VRS). The VRS assumption provides the measurement of pure technical efficiency (PTE), which is the measurement of technical efficiency devoid of the scale efficiency (SE) effects. If there appears to be a difference between the technical efficiency (TE) and PTE scores of a particular DMU, then it indicates the existence to scale inefficiency [11].

DEA begins with a fractional programming formulation. Assume that there are n DMUs to be evaluated. DMU_j consumes x_{ij} amounts of input to produce y_{rj} amounts of output. It is assumed that these inputs, x_{ij}, and outputs, y_{rj}, are non-negative, and each DMU has at least one positive input and output value. The productivity of a DMU can be written as:

$$h_j = \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}}.$$

In this equation, u and v are the weights assigned to each input and output. By using mathematical programming techniques, DEA optimally assigns the weights subject to the following constraints. The weights for each DMU are assigned subject to the constraint that no other DMU has efficiency greater than 1 if it uses the same weights, implying that efficient DMUs will have a ratio value of 1. The objective function of DMU is the ratio of the total weighted output divided by the total weighted input:

$$\max h_0(u,v) = \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}},$$

subject to

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1, j = 1,2 \dots, j_0, \dots, n,$$

$$u_r \geq 0, r = 1,2, \dots, s,$$

$$v_i \geq 0, i = 1,2, \dots, m,$$

where h_0 is the technical efficiency of DMU_0 to be estimated, u_r and v_i are weights to be optimized, y_{rj} is the observed amount of output of the r^{th} type for the j^{th} DMU, x_{ij} is the observed amount of input of the i^{th} type for the j^{th} DMU, r indicates the s different outputs, i denotes the m different inputs and j indicates the n different DMU_s.

2.2 Malmquist index

The Malmquist index [8] evaluates efficiency change over time. The MI, based on DEA models, is one of the prominent indexes for measuring the relative productivity change of DMUs in multiple time periods. This index breaks down into various components. The index provides a useful way of distinguishing between changes in technical efficiency, pure technical efficiency, scale, total factor productivity (TFPC) and shifts in the efficiency frontier (technological change) over time. This index is the geometric mean of two TFPC indices, one evaluated with respect to the technology (efficiency frontier) in the current period t and the other with respect to the technology in the base period s [5]. One extension with DEA is to apply MI to panel data to estimate changes in technical efficiency, technological progress and total factor productivity.

The original idea of the MI was proposed by [8] who suggested comparing the input of a firm at two different points of time in terms of the maximum factor by which the input in one period could be decreased such that the firm could still produce the same output level of the other time period. [2] extended the original Malmquist input index and introduced the first type of the Malmquist index, and then [6] showed that the Malmquist index can be calculated using a nonparametric DEA-like approach, given that suitable panel data are available and they applied DEA for measuring the Malmquist index. They assumed constant returns to scale and identified the technological change and the change of technical efficiency as two components of the productivity changes over time. Next, [7] considered variable return to scale and offered an extended decomposition of the Malmquist index with another important factor capturing change in scale efficiency.

Following [7] we use DEA to construct an input based MI between period t (the base period) and period s:

$$M_I(y^s, x^s, y^t, x^t) = \left[\frac{D_I^t(y^s, x^s)}{D_I^t(y^t, x^t)} x \frac{D_I^s(y^s, x^s)}{D_I^s(y^t, x^t)}\right]^{\frac{1}{2}},$$

where $M_I(\cdot)$ is the input-oriented MI, $D_I^t(y^s, x^s)$ is the distance function showing a maximal proportional reduction of the observed period *s* inputs under the period *t* technology. The distance function is defined as follows:

$$D_I^t(y^s, x^s) = \min_{\theta,\lambda} \theta$$

subject to

$$y_{is} \leq \lambda Y^t$$

$$\theta x_{is} \geq \lambda X^t$$

$$\lambda_i \geq 0, i = 1, \dots n$$

where θ is a scalar and λ is a vector of constants. The value of θ obtained is the component score of the *i*-th firm. *X* and *Y* are input and output vectors, and the amounts of the i^{th} input consumed and output generated by the DMU_0, are denoted by *x* and *y*.

2.3 Data and selection of variables

The data set used in this study was obtained from the annual reports of commercial banks during the period 2001–2010 and all the data is reported on unconsolidated basis. We analyzed only commercial banks that are operating as independent legal entities. As we have reliable data extracted directly from annual reports we eliminate the risk that incomplete or biased data may distort the estimation results.

In order to conduct the DEA estimation, inputs and outputs need to be defined. In the empirical literature four main approaches have been developed to define the input-output relationship in financial institution behavior (intermediation, production, asset and profit approach). We adopt intermediation approach which assumes that the banks' main aim is to transform liabilities (deposits) into loans (assets). Consistently with this approach, we assume that banks use the two inputs. The bank collects deposits to transform them, using labor, in loans. We employed two inputs (labor and deposits), and two outputs (loans and net interest income). We measure labor by the total personnel costs covering wages and all associated expenses and deposits by the sum of demand and time deposits from customers, interbank deposits and sources obtained by bonds issued. Loans are measured by the net value of loans to customers and other financial institutions and net interest income as the difference between interest incomes and interest expenses.

3 Empirical analysis and results

DEA can be used to estimate efficiency under the assumptions of constant and variable returns to scale. For empirical analysis we use DEAP 2.1 software (A Data Envelopment Analysis Computer Program), which was written by Tim Coelli. The DEA method is suitable in the banking sector because it can easily handle multiple inputs-outputs producers such as banks and it does not require the specification of an explicit functional form for the production frontier or an explicit statistical distribution for the inefficiency terms unlike the econometric methods [10]. The banking efficiency have been estimated using the DEA models, input-oriented model with constant returns to scale and input-oriented model with variable returns to scale. The reason for the using of both techniques is the fact that the assumption of constant returns of scale is accepted only in the event that all production units are operating at optimum size. This assumption, however, in practice it is impossible to fill, so in order to solve this problem we calculate also with variable returns of scale.

The results of the DEA efficiency scores based on constant returns to scale are presented in Table 1. Dresdner Bank has the efficiency score of 100% in 2001–2003 and then BAWAG has the average efficiency score 84%. Volksbank, GE Money Bank and JT bank are considered to be efficient with the efficiency scores of 100%, implying that it had produced its output on the efficiency frontier in most analyzed years. HVB has the efficiency scores of 100% in 2002, 2004 and 2006. IC bank has the average efficiency 97% and Banco Popolare has the average efficiency 96%. eBanka, ČSOB, Česká spořitelna and Komerční banka have the average efficiency score less than 50%.

It can be concluded that the largest banks in the market appeared to be least efficient. Considerable inefficiency was also revealed in mid-sized banks that are building up the market position and using aggressive business strategies. One of the advantages of DEA is that the model identifies sources of lower efficiency. In the Czech banking sector, the main source of inefficiency is the excess of client deposits managed by banks. To a lesser degree, low weight in calculation process was often assigned to net interest income.

DMU	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	Mean
CSOB	0.70	0.28	0.79	0.40	0.68	0.02	0.41	0.29	0.01	0.08	0.37
CS	0.38	0.07	0.83	0.74	0.44	0.31	0.22	0.22	0.16	0.22	0.36
KB	0.23	0.17	0.94	0.32	0.18	0.12	0.08	0.36	0.32	0.23	0.29
HVB	0.87	1.00	0.99	1.00	0.85	1.00					0.95
UNIC							0.73	0.88	0.76	0.28	0.66
ZIBA	0.63	0.68	0.67	0.75	0.79	0.90					0.73
GEM	0.75	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.87	0.28	0.89
RB	0.76	0.72	0.67	0.86	0.75	0.88	0.88	0.82	1.00	0.42	0.78
IC	1.00	1.00	0.95	1.00	0.95	0.92					0.97
POPO							0.95	0.89	1.00	1.00	0.96
JTB	0.66	0.85	1.00	1.00	1.00	1.00	0.89	0.92	1.00	0.78	0.91
DRESD	1.00	1.00	1.00								1.00
BAWA				0.65	0.71	1.00	1.00				0.84
LBBW								1.00	1.00	0.64	0.88
PMB	0.83	0.95									0.89
PPF			0.93	0.94	0.58	0.64	1.00	1.00	0.55	1.00	0.83
VOLKS	1.00	1.00	1.00	1.00	1.00	1.00	0.98	1.00	0.55	0.48	0.90
CITI	1.00	0.73	0.75	0.63	0.53	0.43	0.37				0.63
EBAN	0.30	0.20	0.43	0.63	0.61	0.50	0.44				0.44
Mean	0.72	0.69	0.85	0.78	0.72	0.69	0.69	0.76	0.66	0.49	

Table 1 Efficiency of the Czech banks in CCR model

Table 2 reports efficiency scores obtained relative considering variable returns to scale for each year. HVB bank, IC bank, Banco Popolare and Dresdner bank and then LBBW bank have efficiency score of 100% in all analyzed years. UniCredit bank, GE Money Bank, Raiffeisenbank, JT Bank, BAWAG bank and Volksbank are considered to be fully efficient with the efficiency scores of 100% over all analyzed years. Česká spořitelna and Komerční banka have the efficiency score of 100% in 2010. ČSOB, Česká spořitelna, Komerční banka and eBanka have the average efficiency score less than 50%. Efficiency scores of almost all large banks improve when the assumption of variable returns of scale built in BCC model is used.

DMU	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	Mean
CSOB	1.00	0.34	0.79	0.41	0.68	0.03	0.42	0.38	0.05	0.09	0.42
CS	0.55	0.08	0.83	0.76	0.45	0.33	0.24	0.28	0.18	1.00	0.47
KB	0.29	0.17	0.94	0.34	0.20	0.14	1.00	0.44	0.42	1.00	0.49
HVB	1.00	1.00	1.00	1.00	1.00	1.00					1.00
UNIC							1.00	1.00	1.00	0.31	0.83
ZIBA	0.67	0.68	0.67	0.76	0.86	0.92					0.76
GEM	0.75	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.33	0.91
RB	0.91	0.82	0.73	0.91	0.85	0.91	1.00	0.96	1.00	0.82	0.89
IC	1.00	1.00	1.00	1.00	1.00	1.00					1.00
POPO							1.00	1.00	1.00	1.00	1.00
JTB	0.72	1.00	1.00	1.00	1.00	1.00	0.93	0.93	1.00	0.84	0.94
DRESD	1.00	1.00	1.00								1.00
BAWA				0.65	0.75	1.00	1.00				0.85
LBBW								1.00	1.00	1.00	1.00
PMB	0.95	1.00									0.97
PPF			0.97	0.95	0.60	0.65	1.00	1.00	0.59	1.00	0.84
VOLKS	1.00	1.00	1.00	1.00	1.00	1.00	0.98	1.00	0.91	0.72	0.96
CITI	1.00	0.79	0.80	0.64	0.55	0.43	0.39				0.66
EBAN	0.44	0.22	0.44	0.63	0.61	0.50	0.44				0.47
Mean	0.81	0.72	0.87	0.79	0.75	0.71	0.80	0.82	0.74	0.74	

Table 2 Efficiency of the Czech banks in BCC model

In the Czech banking sector, there are several common features, which are valid over the whole analyzed period. The DEA model indicates that the reason of lower efficiency in the Czech banking industry is persistently low efficiency of utilization of fixed assets. Banks hold excessive fixed assets mainly in the form of buildings. Generally, during the period 2001–2010, the average efficiency calculated using the CRS ranges from 49 to 85% and the average efficiency computed using the VRS ranges from 71 to 87%. It shows that the Czech banks are in average considered to be efficient with only marginal changes over time.

The results of the CCR model and the BCC model show that the model with VRS achieves higher degree of the efficiency than the model with the CRS. Number of efficient banks is higher in the model with VRS. [11] argued this is because the BCC model decomposes inefficiency of production units into two components: the pure technical inefficiency and the inefficiency to scale. The values of efficiency computed by the BCC model reach higher values than efficiency computed by the CCR model by eliminating the part of the inefficiency that is caused by a lack of size of production units.

The large volume of information derived from DEA may be difficult to summarize and evaluate. Therefore, it is often helpful to break down the information using the Malmquist index. We calculate MI from the DEA scores between adjacent periods. The application of the MI is also conducted in DEAP 2.1 software. The Malmquist change indices are computed using DEA. The indices measure TFPC for sampled banks in adjacent year during the period of 2001/2002 and 2009/2010. Its decomposition into technical efficiency change (TEC), technological change (TCC), pure technical efficiency change (PTEC) and scale efficiency change (SEC) components are also derived. The VRS or CRS option has no influence on the MI because both are used to calculate the various distances used to construct the Malmquist indices.

One important point is that the calculation of the Malmquist TFPC index requires strictly balanced panel data. That means all DMUs must be observed in all time periods. But this study has an unbalanced panel. One way of dealing with the unbalanced panel data is to find out as many DMUs as possible that are observed in the longest common periods. This method has been the most commonly used by many researchers, but it has one existing shortcoming of this method – greatly reducing the research sample size. We use panel data of 11 Czech banks (with regard to mergers and acquisitions of banks). Table 3 presents the results of the average Malmquist indices in the analyzed period.

	2002	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**Mean**
TEC	0.763	0.757	1.664	1.089	0.973	0.995	0.956	0.976	1.154	1.01
TCC	1.142	2.902	0.509	0.77	0.963	0.998	1.057	0.923	0.745	0.944
PTEC	0.91	0.995	1.049	0.982	0.971	1.007	0.989	0.99	1.084	0.997
SEC	0.839	0.761	1.586	1.109	1.002	0.988	0.967	0.986	1.064	1.013
TFPC	0.872	1.439	0.847	0.839	0.937	0.993	1.011	0.901	0.859	0.953

Table 3 Malmquist index of the sample banks

The average efficiency change reaches the -4.7 %. This negative efficiency change can be dichotomized into its catch-up and frontier-shift components. The mean value of TEC (catch-up or recovery component) registered 1.01, or under 1.00 indicating progress or positive efficiency change. The catch-up effect is comprised of pure and scale efficiency changes. Pure efficiency change represents core efficiency due to improved operations and management while scale efficiency change is associated with returns to scale effects. PTEC reached below 1 on average for the period suggesting regress in terms of operations and management, and SEC reached value under 1 showing the positive scale economies effects. Technological change or frontier-shift represents the innovation in the banking system that has been developed, adapted or absorbed by the players. Technological change is average 0.944. The results of the Malmquist index reached an annual average negative growth of -4.7%. This negative change can be dichotomized into efficiency change and technological change. Technological change reached an average annual negative growth of -5.6% and technical efficiency change improved average by 1%.

	CSOB	**CS**	**KB**	**UNIC**	**GEM**	**RB**	**POPO**	**JTB**	**LBBW**	**PPF**	**VOLKS**
TEC	1.024	1.031	1.027	1.034	1.01	1.016	1	1	0.992	1	0.976
TCC	0.91	0.991	0.888	0.95	0.964	0.937	0.929	1.042	0.933	0.929	0.918
PTEC	1	0.99	0.973	1.016	1	1	1	1	0.984	1	1
SEC	1.024	1.041	1.056	1.017	1.01	1.016	1	1	1.009	1	0.976
TFPC	0.932	1.021	0.912	0.982	0.974	0.952	0.929	1.042	0.926	0.929	0.896

Table 4 Malmquist index of the individual banks

The individual banks' Malmquist indices are given in Table 4. The average efficiency change achieved the negative growth in most banks (besides Česká spořitelna and JT banka) indicating that analyzed banks registered negative efficiency growth in the Czech banks. We decomposed the efficiency change into the catch-up and frontier-shift effects and found that the catch-up effect was primarily accountable for the productivity growth rather than the frontier-shift effect, suggesting that the industry has lacked innovation or technological progress in the past 10 years. Technological efficiency change register below 1.00 in all analyzed banks. It means that technology has a negative effect on the total efficiency change. The values of catch-up registered values under 1.00 in most banks (besides LBBW), which indicate the progress or positive efficiency change. PTEC and SEC reached values under 1 in the most Czech banks, which suggest progress in terms of operations and management, and positive scale economies effects.

4 Conclusion

The aim of this paper was to estimate the technical efficiency and the efficiency change in the Czech commercial banks during the period 2001–2010. For the estimation we applied the Data Envelopment Analysis and the Malmquist index on the data of the Czech banks. In the paper it was found that the average efficiency computed under the assumption of constant returns to scale ranges from 49 to 85 % and the average efficiency estimated under the assumption of variable returns to scale ranges from 71 to 87 %. We found that the efficiency scores from the BCC model reached higher values than efficiency scores from the CCR model by eliminating the part of the inefficiency that is caused by an inappropriate size of production units. Large banks in the market appeared to be inefficient. In the Czech banking sector, the main source of inefficiency is the excess of client deposits managed by banks. The results of the Malmquist index reached an annual average negative growth of -4.7%. It was found that the catch-up effect was primarily accountable for the productivity growth rather than the frontier-shift effect, suggesting that the industry has lacked innovation or technological progress in the past 10 years. The average efficiency change achieved the negative growth in most banks (besides Česká spořitelna and JT banka). This negative change can be dichotomized into efficiency change and technological change. Technological change reached an average annual negative growth of -5.6% and technical efficiency change improved average by 1%.

Acknowledgements

Research behind this paper was supported by the Student Grant Competition of Silesian University within the project SGS 25/2010 'Financial integration in the EU and its effect on corporate sector'.

References

[1] Banker, R. D., Charnes, A., and Cooper, W. W.: Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science* **30** (1984), 1078–1092.
[2] Caves, D. C., Christensen, L. R., Dievert, W. E.: The economic theory of index number and the measurement of input, output, and productivity. *Econometrica* **50** (1982), 1393–1414.
[3] Charnes, A., Cooper, W. W., and Rhodes, E.: Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research* **2** (1978), 429–444.
[4] Charnes A., Cooper, W. W., Lewin, A. Y., and Seiford L. M.: *Data Envelopment Analysis: Theory, Methodology and Applications*. Springer-Verlag, New York, 1995.
[5] Coelli, T. J., Prasada Rao, D. S. and Battese, G. E.: *An Introduction to Efficiency and Productivity Analysis*. Kluwer Academic Publishers, Boston, 1998.
[6] Fare, R., Grosskopf, S., Lindgren, B., and Roose, P.: Productivity change in Swedish analysis, Pharmacies 1980–1989: A nonparametric Malmquist approach. *Journal of Productivity* **3** (1992), 85–102.
[7] Fare, R., Grosskopf, S., Norris, M., and Zhang, A.: Productivity growth, technical progress, and efficiency changes in industrial country. *American Economic Review* **84** (1994), 66–83.
[8] Malmquist, S.: Index numbers and indifference surfaces. *Trabajos de Estadistica* **4** (1953), 209–242.
[9] Seiford, L. M., and Thrall, R. M.: Recent developments in DEA: the mathematical programming approach to frontier analysis. *Journal of Econometrics* **46** (1990), 7–38.
[10] Singh, G., Singh, P., and Munisamy, S.: A Cross Country Comparison of Banking Efficiency: Asia Pacific Banks. *International Review of Business Research Papers* **4** (2008), 73–95.
[11] Sufian, F.: The Efficiency of Islamic Banking Industry: a non-parametric analysis with non-discretionary input variable. *Islamic Economic Studies* **14** (2007), 53–78.
[12] Vincová, K.: Neefektívnosť z rozsahu v bankovom sektore: Komparácia slovenského a českého bankového sektora. In: *International Conference Proceedings: National and Regional Economics VI*. Ekonomická Fakulta, Technická Univerzita v Košiciach, Košice, 2006, 440–445.

Impact of the global financial crisis on stock market volatility: Evidence from Central European stock market

Petr Seďa[1]

Abstract. Volatility measuring is an important task in financial markets, and it has held the attention of academics and practitioners over the last two decades. This paper deals with the impact of the global financial crisis on Central European stock market volatility represented by the Czech and Polish stock markets. Therefore, fluctuations and volatility in these markets before, during and after the crisis were analyzed. This paper tries to identify the length of the global financial crisis, estimate the potential risk in the stock market during financial turmoil and analyze the characteristics of the risk. For a comprehensive analysis, several sophisticated models of quantitative financial analysis were adopted. We especially worked with Jump-Diffusion GARCH model considering heteroskedasticity which allow greater accuracy than simple GARCH type volatility models.

Keywords: Global Financial Crisis, Central European Stock Market, Heteroskedasticity, Annualized Return, Volatility, Jump-Diffusion GARCH model.

JEL Classification: C 32, C 52, C 58, E 32

1 Introduction

Financial markets, due to their key role in the economic positions of countries, have been studied from different points of view. In this regard, one key aspect of the stock markets that has attracted much attention in financial literature is the analysis of the stock returns and its volatility. Ups and downs in prices are quite natural in stock market. Volatility is a symptom of a highly liquid stock market. Investors interpret a raise in stock market volatility as an increase in the risk of investment and consequently they shift their funds to less risky assets.

Volatility modeling and especially volatility dynamics are central to many issues in financial markets including derivative prices, leverage ratios, credit spreads, and portfolio decisions. In times of low market volatility it is relatively straightforward to measure volatility and understand volatility dynamics. At other times, financial markets are affected by severe disruptions which may be largely isolated events like the market crash of 1987, may be a series of events such as the Russian default and Long Term Capital Management Fund Crisis of 1998 or the global financial crisis in 2008-2009. During such periods, apparent spikes in volatility and large movements in asset prices complicate estimation of volatility and volatility dynamics. These crises dramatically influenced the market volatility and diversification opportunities for foreign investors.

The global financial crisis that still affects countless countries originated in the USA as a financial meltdown that materialized as the housing market bubble burst. Various financial institutions were hit by the crisis, which by then was no longer merely a problem of the USA. Among the top five global investment banks of the US, Merrill Lynch was sold to Bank of America, and Lehman Brothers filed for bankruptcy. The crisis began to affect the real economy. Czech Republic, an export-oriented economy with high liquidity and substantial reliance on foreign capital, was no exception. The fund withdrawal led by foreign investors in the Czech securities market exacerbated volatility in the securities market. Polish stock market can be described in a similar way.

Recently, several authors have investigated the volatility of Central and Eastern European stock markets; see Baruník et al. [2], Popelka [12] and Seďa [12] found that significant autocorrelation, high volatility persistence, significant asymmetry, lack of relationship between the stock market volatility and the expected return and non-normality of the return distribution are basic characteristics of the stock market volatility in transition countries. Nevertheless, we can identify dealing with the influence of market crises over Central European stock market volatility in the literature quite rarely.

This paper deals with the impact of the subprime mortgage meltdown born in USA on the volatility of Czech and Polish security market as it spiraled into a global financial crisis. Policy makers and economic players are interested in whether the state and length of the global financial crisis could be measured accurately in stochastic terms. Moreover, the structural characteristics of risk to stock prices fluctuation are analyzed. For a comprehensive analysis, some sophisticated models of quantitative financial analysis were adopted. To identify the impact

[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.

of the global financial crisis on Central European stock markets, rapid fluctuations and volatility in these markets during the crisis were analysed.

The studies conducted in this paper can be summarized as follows: First, descriptive statistics of Czech and Polish stock markets and time series features will be computed. Second, annualized conditional volatility of mentioned returns will be estimated by the standard AR (1) - GARCH (1, 1) - GED model. Last, time-varying volatility of stock price indexes was analyzed in the pre-crisis, crisis and post-crisis periods using Jump-Diffusion GARCH model to identify jump risk in volatility.

Innumerable expert studies have been provided to identify the volatility of stock prices, and the structural characteristics of risk. The adoption of stochastic volatility to let volatility change stochastically and the inclusion of jump risk to reflect the risk of rapid fluctuation in the market in a model can be seen in Merton [11], Hull and White [10], Heston [9], and Chang [6]. Papers that have dealt with the characteristics of basic time series such as heteroscedasticity of daily financial time series and jump are the ARCH model by Engle [8], Ball and Torous [1] and the Jump-Diffusion model of Chang and Kim [4]. The aim of this paper is to identify and estimate the potential jump risk in the stock market volatility comparing financial crisis period with the time of normalcy and using data from Czech and Polish stock markets.

2 Jump-diffusion GARCH model

The seminal papers of Engle [8] and Bollerslev [3] GARCH or generalized autoregressive conditional heteroskedasticity models have become a standard tool in modeling the conditional variances of the returns from financial time series data. The popularity of these models lies in their compatibility with major stylized facts for asset returns, the existence of efficient statistical methods for estimating model parameters, and the availability of useful volatility forecasts.

The stock index returns may have jump risk at the same time which is caused mainly by heteroskedasticity and rapid market fluctuation. In this chapter we will take a look on how to consider heteroskedasticity and jump risk term in the model in order to identify the process of forming appropriate return and risk premium in such case. Normally, the jump risk model taking the form of discrete data and considering heteroskedasticity in GARCH (1, 1) may be according to Chang et al. [7] defined as follows:

$$r_t = \alpha_0 + \alpha_1 r_{t-1} + k\sigma_t + \varepsilon_r + \sum_{j=0}^{q_t} v_{jt}, \tag{1}$$

$$\varepsilon_t = \xi_t \sigma_t, \tag{2}$$

$$\xi_t \sim N(0,1), \tag{3}$$

$$q_t \sim e^{-\lambda} \frac{\lambda^j}{j!}, \tag{4}$$

$$v_t \sim N\left(0, v^2\right), \tag{5}$$

$$\sigma_t^2 = \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 \sigma_{t-1}^2, \tag{6}$$

where r_t is the return of stock index estimated here, which shows the volume of information inflow in the market. It is determined by the independently and log-normally distributed jump v_{jt} and the Poisson distributed random variable q_t. The mean of q_t is recognized as λ the intensity of jump. The Poisson event causes a heteroskedastic jump with each jump size of $exp\left(v_j\right), j = 1,2,...,q_t$ in stock prices return. Therefore, v_{jt} is assumed as a random variable with a mean of 0, variance of v^2, as an *i.i.d.* normal distribution. Parameter k represents time-varying risk premium.

The GARCH heteroskedasticity is defined by equation (6). $\beta_0, \beta_1, \beta_2$ are constants satisfying the conditions of $\beta_0 > 0, \beta_1 + \beta_2 < 1$. If $\sum_{j=0}^{q_t} v_{jt}$, the Poisson jump risk term, is insignificant, the model will be equal to the GARCH model of Bollerslev [3]. Equations (1) to (6) that include unobserved state variable can be gained by using the Kalman filter model after converting into state-space models. Detailed estimation can be according Kim and Chang [5] referred as follows:

$$r_t = r(g_t) + G(g_t)\alpha_t + \varepsilon_t, \tag{7}$$

$$\alpha_t = F_t \alpha_{t-1} + R\eta_t. \tag{8}$$

In the diffusion-jump model that considers GARCH volatility, the return volatility of index is h_t which can be calculated as the weighted average of the diffusion and jump parts:

$$h_t = \sum_{j=0}^{J} \left(\sigma_t^2 + q_t \bullet v^2\right) Pr[q_t = j \mid \psi_t]. \tag{9}$$

The GARCH conditional volatility $\sigma_{t+k,t}^2$ of point $t+k$ considered at point t of the diffusion part in equation (9) can be denoted for repetitive calculation:

$$\sigma_{t+k,t}^2 = \beta_0 + \beta_1 \left(\sum_{j=0}^{J} \left(\sigma_{t+k-1,t}^2 + q_t \bullet v^2\right) \varpi_j^{k-1} \Pr[q_t = j \mid \psi_t] \right) + \sigma_{t+k-1,t}^2. \tag{10}$$

3 Empirical application

In this chapter a data description and estimation results will be presented. The studies will be summarized as described in the Introduction section.

3.1 Data

Empirical analysis is performed on daily data of PX and WIG20 indexes in period from 2004 till 2012, it includes total of 2225 observations. This period was chosen purposely, to investigate changes of the Czech and Polish equity markets volatility during time with a special emphasis on the resolution of behavior in the time before, during and after the global financial crisis in 2008-2009. We have more than 8 years long time series of the closing rates of PX and WIG20 indexes. Those time series were obtained from www.pse.cz and www.wse.com.pl.

The returns r_t at time t were defined as the logarithm of PX and WIG20 indices p, that is, $r_t = \log(p_t - p_{t-1})$. Visual inspection of the plot of daily values and returns series of both indices proved very useful, for details see Figure 1 and Figure 2. As it has been already empirically confirmed, crises are not devoted to developed markets only. Emerging markets includes Czech Republic and Poland isn't excluded from this rule and may face such instability sometime. Following the spread of bad news about U.S financial crisis the Central European equity markets, Czech and Polish ones included, we have seen a more than 60 percent decline of both selected indexes in 2008, see Figure 1. This happened primarily due to the withdrawal by foreign portfolio investors between September and December 2008 and its psychological impact on national investors.

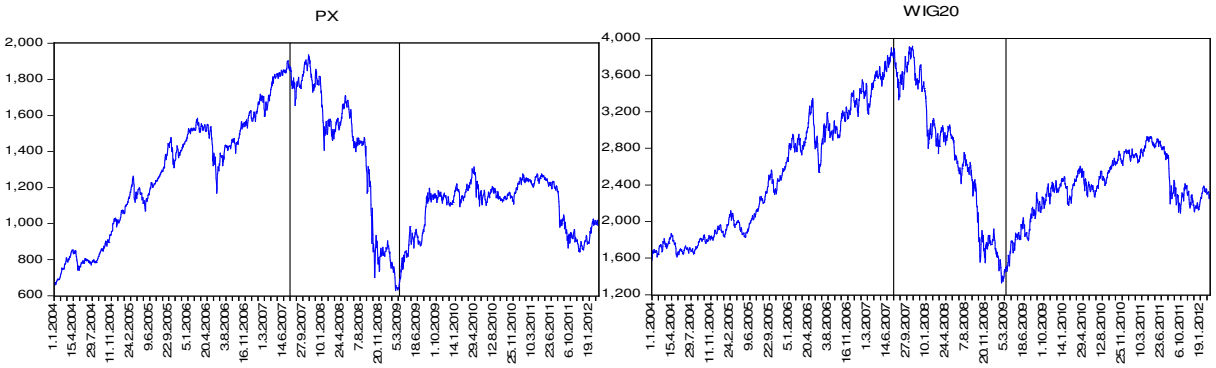


Figure 1 PX and WIG20 values (2004-2012)

It can be seen that from Figure 2 that return fluctuates around mean value that is close to zero. Volatility is low for certain time periods and high for other periods. The movements are in the positive and negative territory and larger fluctuations tend to cluster together separated by periods of relative calm. From 2004 to early 2007, the financial markets had been very calm in general. The volatility of PX and WIG20 indexes was highest in 2008. Thus Figure 2 show volatility clustering where large returns tend to be followed by small returns leading to continuous periods of volatility and stability. The market volatility, as measured by the PX and WIG20 volatility have been below long-term averages. However, the global financial crisis of 2008 changed this: most asset

classes experienced significant pullbacks, the correlation between asset classes increased significantly and the markets have become extremely volatile.

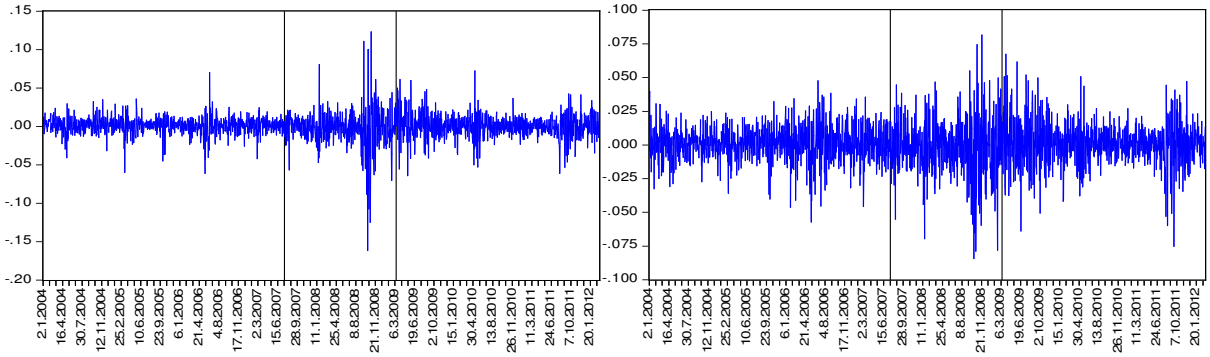


Figure 2 PX and WIG20 returns (2004-2012)

Volatility clustering implies a strong autocorrelation in squared return. Since the volatility was highest in 2008 when the values of both indexes reached the minimum values in investigated period we divided the basic period 2004-2012 into three testing periods. First, pre-crisis period, was defined from January 2004 to the end of June 2007, the second, crisis period, started at the beginning of July 2007 and finished by March 2009 and the last, post-crisis period, was defined from April 2009 to the middle of March 2012. Our goal is to investigate and compare the behaviour of volatility in all the periods.

3.2 Estimation results

Table 1 shows several descriptive statistics and time series features of the PX and WIG20 indexes returns of the normalcy, the financial crisis and post crisis periods.

	PX			WIG20		
	Pre-crisis	**Crisis**	**Post-crisis**	**Pre-crisis**	**Crisis**	**Post-crisis**
Mean	0,0011	-0,0025	0,0006	0,0009	-0,0022	0,0006
Maximum	0,0705	0,1236	0,0725	0,0475	0,0815	0,0672
Minimum	-0,0613	-0,1619	-0,0644	-0,0573	-0,0844	-0,0754
St. Deviation	0,0107	0,0251	0,0152	0,0127	0,0218	0,0158
Skewness	-0,5572	-0,3939	-0,0439	-0,3235	-0,2187	-0,0234
Kurtosis	8,6038	12,1049	5,7086	4,5116	4,7066	5,4142
Jarque-Bera test	1254,09	1492,78	239,91	103,86	55,48	190,48
Probability	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Observation	922	429	784	922	429	784

Table 2 Descriptive statistics of PX and WIG20 indexes

According to Table 1, the daily PX and WIG20 returns show mostly leptokurtic distribution with a heavy tail, instead of normal distribution. The sample statistics of analysed time series data indicate that it is desirable to consider heteroscedasticity and jump risk when estimating the volatility of the Czech and Polish markets. As expected, the volatility of stock price is much larger during the financial crisis than in time of normalcy.

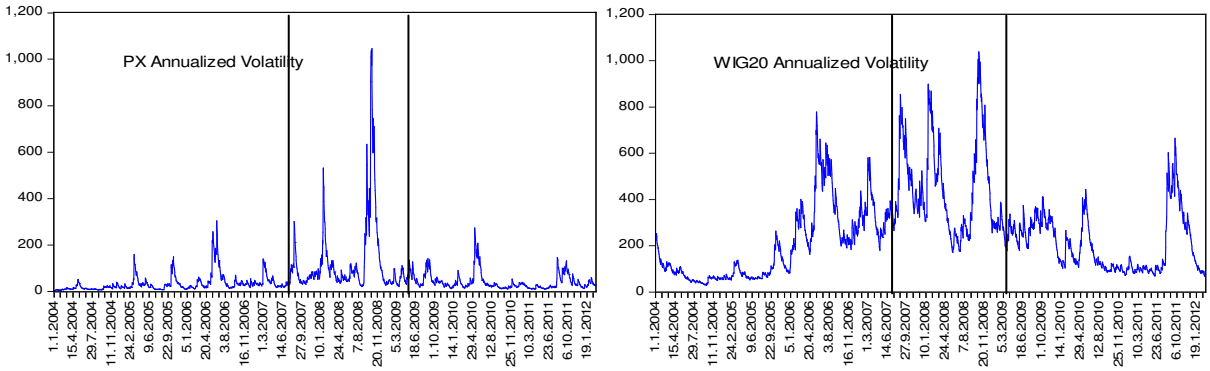


Figure 3 Conditional volatility of PX and WIG20 indexes (2004-2012)

Figure 3 shows annualized conditional volatility of PX and WIG20 indexes return estimated by the standard AR (1) -GARCH (1, 1) - GED model. Annualized conditional volatility can be calculated from the daily volatility multiplied by square root of trading days. According Figure 3, conditional volatility of WIG20 index is larger than PX index volatility in general. In particular, in the case of WIG20 return volatility, risk has doubled compared to PX return volatility on average. Conditional volatility of both indexes grows significantly during the global financial turmoil.

Parameter	Pre-crisis		Crisis		Post-crisis	
	Estimate	***t*-value**	**Estimate**	***t*-value**	**Estimate**	***t*-value**
α_0	0,0016	5,49	-0,0005	-0,63	0,002	0,59
α_1	0,0264	1,72	0,0140	0,26	0,0006	0,16
β_0	0,0001	2,64	0,0001	1,72	0,0003	2,11
β_1	0,1076	4,79	0,1856	3,38	0,1156	4,75
β_2	0,8341	21,74	0,8007	15,79	0,8697	36,13
λ	0,0005	1,15	0,2165	2,49	0,0451	1,87
v	0,0023	0,02	0,0466	1,24	0,0071	0,05
k	-0,0579	-1,41	0,0142	1,16	-0,0396	-1,29

Table 2 Jump-Diffusion GARCH (1, 1) model with heteroskedasticity for PX index

Parameter	Pre-crisis		Crisis		Post-crisis	
	Estimate	***t*-value**	**Estimate**	***t*-value**	**Estimate**	***t*-value**
α_0	0,0011	2,90	-0,0016	-1,79	0,0005	1,16
α_1	0,0090	0,27	-0,0271	-0,51	-0,0205	-0,57
β_0	0,0000	1,68	0,0000	1,49	0,0000	1,55
β_1	0,0441	2,91	0,0913	2,94	0,0676	4,02
β_2	0,9371	43,95	0,8894	24,07	0,9236	51,30
λ	0,1314	1,56	0,2729	3,12	0,0981	2,06
v	0,0001	0,01	0,0006	0,47	0,0004	0,26
k	-0,0705	-1,13	0,0638	1,86	-0,0679	-1,11

Table 3 Jump-Diffusion GARCH (1, 1) model with heteroskedasticity for WIG20 index

Table 2 and Table 3 show the estimations for PX and WIG20 indexes by period. There are the results of the Jump-Diffusion GARCH model with heteroscedasticity proposed by equations (1) to (6) and estimated using Maximum Likelihood method. According to the Jump-Diffusion GARCH model considering heteroscedasticity, no statistically significant jump behaviour was observed in the Czech and Polish stock market in the pre-crises period, between January 2004 and July, 2007. On the other hand, during the global financial crisis, from July 2007 to March 2009, the jump risk with relatively high statistical significance occurred in both stock markets around every five days (Czech market) or three-four days (Polish market). Therefore, during the financial turmoil, jump has occurred more frequently in the Polish stock market than in the Czech market. In the post-crisis period the behaviour of the both investigated markets again tended to normalcy. In other words, there was observed statistically significant jumps in volatility only in Polish market around every 10 days.

According to the Jump-Diffusion GARCH model with heteroscedasticity projected during the study, $\beta_1 + \beta_2$ that denote the persistence of variance process in the Czech stock market shows 0.9101 in pre-crisis period and surprisingly 0.9863 during the crisis. In addition, the ARCH factor was larger during the financial crisis than in time of normalcy. Meanwhile, the persistence parameter in the Polish market was 0.9812 during pre-crisis period but 0.9807 during the crisis. Also in the case of the Polish market, the ARCH factor was larger in crisis than in normalcy. Variance process was less persistent during the crisis than in normalcy. In addition, it is noteworthy that the time-varying risk premium in both markets was negative in normalcy but positive in the time of crisis.

It seems that conditional volatility estimated in the jump model shows that the size of volatility in stock price during the financial crisis differs from that projected through the AR (1) - GARCH (1, 1) - GED model. This may prove the necessity of using a more precise model in measuring risk during a crisis.

4 Summary and conclusions

The stock market in Central Europe shows relatively high dependence on foreign capital and various Polish and especially Czech conglomerates and small- and medium sized enterprises are export-oriented. The volatility of PX and WIG20 stock returns has been investigated and modelled using AR (1) – GARCH (1, 1) and Jump-Diffusion GARCH (1, 1) models. This study tried to identify an impact of the global financial crisis, estimate the risk in the stock markets during the financial turmoil, and comprehensively analyse the characteristics of the risk.

Based on results of the Jump-Diffusion GARCH model considering heteroscedasticity, no statistically significant jump behaviour was projected in both stock markets in normalcy, between January 4, 2004 and June 29, 2007. However, during the global financial crisis, from July 2, 2007 to March 31, 2010, jump risk with relatively high statistical significance occurred in both stock markets. In case of the Czech market jump risk appeared every five days, while on the Polish market around every three or four days. As for the post-crisis period, jump with statistical significance appeared around every 10 days on the Polish market only. In summary, jump risk has occurred more frequently in the Polish stock market than in the Czech market, and especially during the financial turmoil.

As a possible extension of this paper, the Markov Switching ARCH model or the SWARCH (k, q) model can be adopted to clearly identify the state of the global financial crisis. Making a stochastic estimation on the point when risk state probability rises in the stock market could provide a stronger empirical result than the case in which the beginning of the crisis is randomly set.

Acknowledgements

The support provided by VSB-TU Ostrava under the SGS project SP2012/172 is kindly announced.

References

[1] Ball, C. A. and Torous, W. N.: On Jump in Common Stock Prices and Their Impact on Call Option Pricing. *Journal of Finance* **40** (1985), 155-173.

[2] Baruník, J., Vácha, L. and Vošvrda, M: Tail Behavior of the Central European Stock Markets during the Financial Crisis, *AUCO Czech Economic Review* **4** (2010), 282-294.

[3] Bollerslev, T.: Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* **31** (1986), 307-327.

[4] Chang, K. H. and Kim, M. J.: Stress Testing of Financial Industries: A Simple New Approach to Joint Stress Testing of Korean Banking, Securities, and Non-Life Insurance Industries. *Asia-Pacific Journal of Financial Studies* **38** (2009), 521-543.

[5] Chang, K. H. and Kim, M. J.: *Financial Econometrics*. Kyungmoonsa, 2002.

[6] Chang, K. H.: Characteristics of Stochastic Volatility in Korean Stock Returns. The Korean Journal of Financial Management **20**, (2003), 213-231.

[7] Chang, K. H., Cho, K. Y. and Hong, M. G.: Stock Volatility, Foreign Exchange Rate Volatility and the Global Financial Crisis, *Journal of Economic Research* **15** (2010), 249-272.

[8] Engle, R. F.: Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* **50** (1982), 987-1007.

[9] Heston, S.: A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Option. *Review of Financial Studies* **6** (1993), 327-344.

[10] Hull, J. and White, A.: The Pricing of Options on Assets with Stochastic Volatilities. Journal of Finance **42** (1987), 281-300.

[11] Merton, R. C.: Theory of Rational Option Pricing, *Bell Journal of Economics and Management Science* **4** (1973), 141-183.

[12] Popelka, J.: *Využití lineárních a nelineárních modelů volatility při analýze českých podílových fondů a akcií.* VŠE, Praha, 2007.

[13] Seďa, P.: Asymmetric Conditional Volatility Modelling: Evidence from Central European Stock Markets. In: *Proceedings of Finanční řízení podniků a finančních institucí´2011*. VŠB – Technical University of Ostrava, Ostrava, 2011, 375-383.

Construction of time schedules using integer goal programming

Veronika Skocdopolova[1]

Abstract. Timetabling at universities is a problem that belongs to difficult problems. It can be solved as a complex integer programming model or the solution can be decomposed into several interrelated stages. This paper presents a sequential integer goal programming model for solving the timetabling problem. At first a three-stage IGP model developed by Al-Husain et al. [2] is presented. This approach decomposes the timetabling problem into three parts, where each stage is optimally solved and the outputs are used as inputs in the next stage. At first teachers are assigned to courses, then courses are assigned to time slots, and finally time slots are assigned to classrooms. This approach enables solving the timetabling problem in a reasonable time. Then a modification of this model adapted to the timetable of summer term 2012 developing process of department of econometrics at University of Economics, Prague is introduced and discussed. Also samples of numerical results of this modification are shown.

Keywords: integer goal programming, timetabling, assignment problem, NP-hard problem

JEL Classification: C61
AMS Classification: 90C29

1 Introduction

Preparing a good time schedule for a university is a difficult task that attracts many researchers. However, in practice the scheduling board is still used more than sophisticated models. The term "good time schedule" is questionable, because there is always someone, who is not satisfied with the given timetable. Nevertheless we can find out some objective criteria that indicate the "good time schedule". However this is not the main point of this article. The objective of this paper is to introduce a goal programming approach to timetabling that was published by Al-Husain et al. [2] and modify their model for preparing the timetable of department of econometrics at University of Economics, Prague.

The thought of constructing sophisticated models for solving the university timetabling problem came out in 70's [7] [8]. There are two main approaches to the timetabling problem. First of them is solving the problem as a one complex model, usually via integer programming [3] [5]. Recall that in general, solving of integer programming models is NP-hard problem (see e.g. [4]). This leads to trying to solve the complex model using various heuristic or metaheuristic methods (see e.g. [1], [6]). These methods bring out solutions that are relatively close to optimal solution in relatively reasonable time. The second approach consists in decomposing the problem into interrelated stages, where outputs of one stage act as inputs in the next stage. In this paper, the second approach is utilized.

The rest of the paper is organized as follows: Part 2 describes a sequential three-stage integer goal programming (IGP) model for faculty-course-time-classroom assignments that was developed by Al-Husain et al. [2]. In the part 3, a modification of Al-Husain's model is presented with application on timetable of department of econometrics at University of Economics, Prague. The last part discusses the results of the modification and outlines the possibilities of future research.

2 A Sequential Three-Stage IGP Model

Al-Husain et al. [2] provide a sequential three-stage integer goal programming model for faculty-course-time-classroom assignments. The scheduling is divided into three stages – the faculty-course assignment stage, the courses-timeslot assignment stage, and timeslot-room assignment stage. "The inputs of every stage are translated into goals and solved according to their order of importance, where goals are given priorities according to their order of importance. The output of every stage, which represents an optimal assignment, is then fed to the next stage to act as an input." [2, p. 158]. The process continues as is shown in the Figure 1.

[1] University of Economics, Prague, Department of Econometrics, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic, skov01@vse.cz.

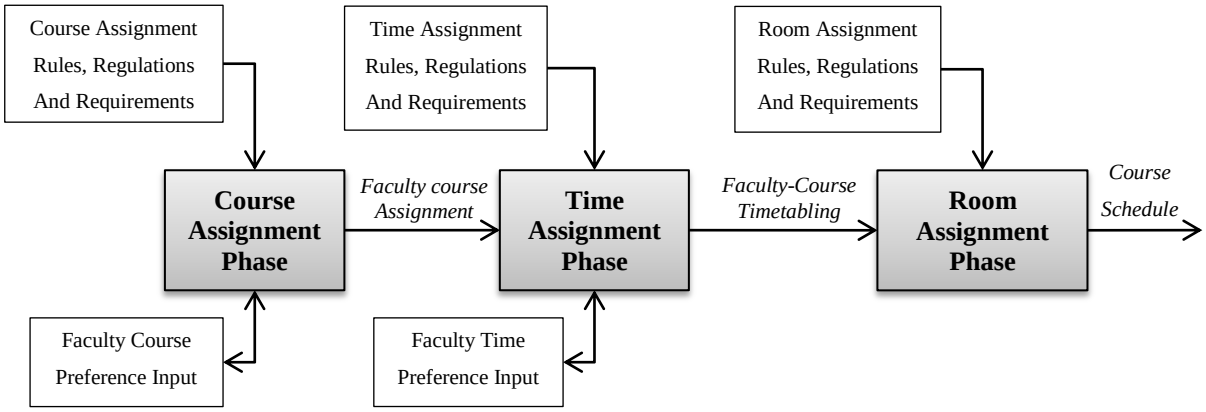


Figure 1 Faculty Course Schedule Block Diagram and Information Flow [2]

In the stage I the courses are assigned to faculty members. The integer GP model has five goals and one strict constraint. The goals are:

1. limit of course loads of each faculty member;
2. number of courses that should be covered by faculty members;
3. each faculty member should take at least one of the College Level Courses (CLC) and
4. at least one of the Major Level Courses (MLC);
5. maximisation of the total preference for each faculty member.

The strict constraint is that any of the faculty members cannot take more than two sections for the same course. In the stage II the courses assigned to faculty members are assigned to time slots. This stage of model has seven goals and three strict constraints. The goals are

1. number of rooms available for each time slot, the number cannot be exceeded;
2. similar CLC assigned to a specific time slot in morning-time cannot exceed 2 sections for the same course;
3. similar CLC assigned to a specific time slot in afternoon-time cannot exceed 1 section for the same course;
4. the MLC should be 4 times more condensed during the morning-time than during the afternoon-time;
5. 60% of courses should be offered during the odd days and 40% during the even days;
6. 70% of courses should be offered during the morning-time and 30% during the afternoon-time;
7. faculty preferences on class times maximisation.

The strict constraints are that sum of sections taught for every faculty in every specific time slot must be at most equal to 1, sum of MLC offered during a specific time slot during same day must equal at most, and sum of time slots for each section for every faculty, every course, and every section must equal 1. The stage III has only one goal and two strict constraints. In this stage the rooms are assigned to the courses. The goal is to locate each previously assigned course to a room of the right size as close as possible to the department that is offering the course. The strict constraints are that each section of a course assigned to a specific faculty and time should be located in one room only, and that each room is assigned to at most one faculty in a specific time period. In the paper [2], the model is applied on scheduling problem at Kuwait University, College of Business Administration.

3 Three-stage IGP model modification

Timetable constructing at the University of Economics starts with sending requirements of each department on classrooms (capacity of required rooms, number of rooms). These requirements are based on past experiences with students' interest in each subject (compulsory or voluntary subject, subject for under graduates or for graduates), on number of students development and on the term which the timetable is prepared for (winter or summer). The requirements are usually a little bit overestimated. On the basis of the requirements, the pedagogical department assigns the classrooms to each department. This assignment is also based on past experiences, this means, when the department A required every winter term 5 classrooms for certain subject ain this term they wanted 12 rooms for this subject, the department will get them only if they give objective reasons for this. And also the pedagogical department knows that the requirements are overestimated, so the departments get less classrooms then they required. In this paper we will deal with constructing the timetable for the department of econometrics for summer term 2012.

The department was assigned by certain classrooms according to its requirements. From the past they estimated the need of course number of each subject. In the next chapters the three-stage model is described.

3.1 Stage I: teacher – course assignment model

In the stage I each course j is assigned to teacher i. The binary decision variable x_{ij} equals 1, if the teacher i is assigned to the course j, and 0 otherwise. At the department there are 32 teachers including PhD students. Every teacher rates each course according to his or her preference of teaching of this course by points SP_{ij}. SP_{ij} gather values from 0 to 5, where 0 means that the teacher i cannot teach course j and 5 points means the teacher i prefer the most to teach course j. Mathematical model of stage I is formulated as follows:

Minimize

$$z = p_1\delta_1^- + p_2\sum_{i=1}^{32}\left(\delta_{2i}^- + \delta_{2i}^+\right) + p_3\sum_{i=1}^{32}\left(\delta_{3i}^- + \delta_{3i}^+\right) \tag{3.1}$$

subject to

$$\sum_{i=1}^{32}\sum_{j=1}^{80} SP_{ij}x_{ij} + \delta_1^- - \delta_1^+ = 400, \tag{3.2}$$

$$x_{ij} \le SP_{ij}, \quad i = 1,2,\ldots,32,\ j = 1,2,\ldots,80, \tag{3.3}$$

$$\sum_{j\in seminar} x_{ij} + \delta_{2i}^- - \delta_{2i}^+ = S_i, \quad i = 1,2,\ldots,32, \tag{3.4}$$

$$\sum_{j\in lecture} x_{ij} + \delta_{3i}^- - \delta_{3i}^+ = L_i, \quad i = 1,2,\ldots,32, \tag{3.5}$$

$$\sum_{i=1}^{32} x_{ij} = 1, \quad j = 1,2,\ldots,80, \tag{3.6}$$

$$\sum_{j\in 4ek202} x_{ij} \le 2, \quad \sum_{j\in 4ek211} x_{ij} \le 2, \quad \sum_{j\in 4ek212} x_{ij} \le 2, \quad \sum_{j\in 4ek213} x_{ij} \le 2, \quad \sum_{j\in 4ek311} x_{ij} \le 2, \quad \sum_{j\in 4ek313} x_{ij} \le 2, \quad i = 1,2,\ldots,32, \tag{3.7}$$

$$\begin{aligned} \delta_{2i}^- &\le 1, \quad i = 1,2,\ldots,32, \\ \delta_{2i}^+ &\le 1, \quad i = 1,2,\ldots,32, \end{aligned} \tag{3.8}$$

$$\begin{aligned} x_{ij} &\in \{0,1\}, \quad i = 1,2,\ldots,32,\ j = 1,2,\ldots,80, \\ \delta_1^-, \delta_1^+, \delta_{2i}^-, \delta_{2i}^+, \delta_{3i}^-, \delta_{3i}^+ &\ge 0, \quad i = 1,2,\ldots,32 \end{aligned} \tag{3.9}$$

where SP_{ij} is the preference of teacher i to teach course j, S_i the maximum number seminar loads for teacher i, L_i the maximum number lecture loads for teacher i. This stage has three goals:

- **Goal 1** – Maximisation of total preference of courses. If each teacher is assigned with course he or she prefers the most, the total preference will be 400 (3.2). This goal has priority p_1 and the objective is to minimize δ_1^-.
- **Goal 2** – Each teacher i should take exactly his or her maximum seminar loads S_i (3.4). This goal has priority p_2 and the objective is to minimize both of deviation variables δ_{2i}^- and δ_{2i}^+ for all teachers i. The deviation variables should be less or equal 1, the teacher should teach one seminar more or less than is his or her maximum load (3.8).
- **Goal 3** – Each teacher i should take exactly his or her maximum lecture loads L_i (3.5). This goal has priority p_3 and the objective is to minimize both of deviation variables δ_{3i}^- and δ_{3i}^+ for all teachers i.

The first hard constraint of this stage (3.3) does not allow teacher i to teach course j, when SP_{ij} equals 0. The second one (3.6) ensures that each course j is assigned to a teacher i. Hard constraints (3.7) ensure, that the teacher i does not teach more than two courses of the same subject. Constraints (3.9) stand for all decision variables are binary and all deviation variables are nonzero.

There are some differences in this stage I model in comparison with Al-Husain's model [2]. In this model binary decision variables are used instead of integer variables. This is because Al-Husain et al. defined the decision variable as number of sections for course *j* assigned to teacher *i*. In our model the courses *j* include all possible sections of all subjects, therefore we use the binary variables that only assign course *j* to teacher *i*. This leads to another difference of the models. Al-Husain et al. use only one set of hard constraints – the decision variables must be less or equal 2 – each teacher cannot take more than two sections of each course. In our model we have to ensure this for subjects with more than two sections by set of constraints (3.7). Last difference is that in this model are omitted two of Al-Husain et al. goals that do not have sense for the department of econometrics.

Results of the stage I is assignment of all courses to teachers according to teachers preferences (Table 1). These results (decision variable x_{ij}) serve as input in the stage II model.

	Course	4EK211C2	4EK211C3	4EK211C4	4EK211C5	4EK211C6	4EK211C7	4EK311C7	4EK313C4	4EK313C5	4EK421C1	4EK421C2	…	4EK425C1	4EK604C1	4EK601P1	4EK602P1
Teacher	1	0	0	0	0	0	0	0	0	0	1	1	…	0	0	0	0
	2	0	0	0	1	0	0	0	0	0	0	0	…	0	0	0	0
	3	0	0	0	0	0	0	0	1	1	0	0	…	0	0	0	0
	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
	30	0	0	0	0	0	1	0	0	0	0	0	…	0	0	0	0
	31	1	0	0	0	0	0	0	0	0	0	0	…	1	1	0	1
	32	0	0	0	0	0	0	1	0	0	0	0	…	0	0	0	0

Table 1 Sample of results of stage I

3.2 Stage II: course – time slot assignment model

Stage II model assigns courses to available time slots. The binary decision variable y_{ijk} equals 1, if the course *j* assigned to the teacher *i* is assigned to time slot *k*, and 0 otherwise. The binary decision variable ajk equals 1, if the course *j* (assigned to the teacher *i*) is assigned to time slot *k*, and 0 otherwise. At the University of Economics, Prague the classwork run in 35 time slots from Monday 7:30 to Friday 19:30. The stage II model can be formulated as follows:

Minimize

$$z = \sum_{k=1}^{35} \delta_{1k}^{+} \tag{3.10}$$

subject to

$$\sum_{k=1}^{35} y_{ijk} \le x_{ij}, \quad i = 1,2,\ldots,32,\ j = 1,2,\ldots,80, \tag{3.11}$$

$$\sum_{j=1}^{80} y_{ijk} \le P_{jk}, \quad i = 1,2,\ldots,32,\ k = 1,2,\ldots,35, \tag{3.12}$$

$$\sum_{i=1}^{32} y_{ijk} = a_{jk}, \quad j = 1,2,\ldots,80,\ k = 1,2,\ldots,35, \tag{3.13}$$

$$\sum_{j=1}^{80} a_{jk} + \delta_{1k}^{-} - \delta_{1k}^{+} = R_k, \quad k = 1,2,\ldots,35, \tag{3.14}$$

$$\sum_{j=1}^{80} y_{ijk} \le 1, \quad i = 1,2,\ldots,32,\ k = 1,2,\ldots,35, \tag{3.15}$$

$$\sum_{k=1}^{35} a_{jk} = 1, \quad j = 1,2,\ldots,80, \tag{3.16}$$

$$\begin{aligned} &y_{ijk} \in \{0,1\}, \quad i = 1,2,\ldots,32,\ j = 1,2,\ldots,80,\ k = 1,2,\ldots,35, \\ &a_{jk} \in \{0,1\}, \quad j = 1,2,\ldots,80,\ k = 1,2,\ldots,35, \\ &\delta_{1k}^{-}, \delta_{1k}^{+} \geq 0, \quad k = 1,2,\ldots,35, \end{aligned} \tag{3.17}$$

where parameter P_{jk} equals 1, if teacher assigned to course j is available to teach the course in time slot k, and 0 otherwise, and R_k is the number of classrooms available in time slot k. The stage II has only one goal (3.14). The total number of courses assigned to a time slot k cannot exceed the number of classrooms available for that time slot. The objective is to minimize the deviation variable δ_{1k}^{+} for all time slots k. The hard constraints (3.11) and (3.13) stand for the cohesion of variables x_{ij}, y_{ijk} and a_{jk}. Constraint (3.12) ensures that teacher assigned for the course j will be available in time slot k. Every teacher i can be assigned most to one time slot k (3.15) and every course j have to be assigned exactly to one time slot k (3.16).

In contrast of the Al-Husain et al.'s 7 goals of stage II model, this model has only one goal. Five of the goals do not have sense for the department and the goal that maximizes the teacher preference on class time was formulated as hard constraint.

3.3 Stage III: course – classroom assignment model

The last stage of the model assigns courses (assigned to time slots) to certain classrooms. The binary decision variable b_{jl} equals 1, if the course j (assigned to the teacher i and to time slot k) is assigned to classroom l, and 0 otherwise. The department has 84 classrooms at disposal. The mathematical model of the stage III can be formulated as follows:

Minimize

$$z = \sum_{l=1}^{84} \delta_{1l}^{+} \tag{3.18}$$

subject to

$$\sum_{l=1}^{84} b_{jl} = 1, \quad j = 1,2,\ldots,80, \tag{3.19}$$

$$\sum_{j=1}^{80} b_{jl} \leq 1, \quad l = 1,2,\ldots,84, \tag{3.20}$$

$$\sum_{j=1}^{80} C_j b_{jl} + \delta_{1l}^{-} - \delta_{1l}^{+} = K_l, \quad l = 1,2,\ldots,84, \tag{3.21}$$

$$\sum_{j=1}^{80} T_j b_{jl} \leq TW_l \sum_{j=1}^{80} b_{jl}, \quad l = 1,2,\ldots,84, \tag{3.22}$$

$$\begin{aligned} &b_{jl} \in \{0,1\}, \quad j = 1,2,\ldots,80,\ l = 1,2,\ldots,84, \\ &\delta_{1l}^{-}, \delta_{1l}^{+} \geq 0, \quad l = 1,2,\ldots,84, \end{aligned} \tag{3.23}$$

where C_j is the required capacity of course j, K_l is the capacity of classroom l, T_j is the time slot assigned to course j and TW_l is the time slot in which is the classroom l available. Stage III has only one goal (3.21). The goal is to locate the course j to classroom with adequate capacity. The objective is to minimize the deviation variable δ_{1l}^{+} for all classrooms l. Each course j has to be assigned exactly to one classroom l (3.19) and each classroom l can be assigned most to one course j. The course j can be assigned to classroom l only in case the classroom l is available in the time slot T_j.

The difference of the stage III model from the Al Husain's model is in omitting the floor level preferences of the classrooms, because the University of Economics, Prague is not too big to make this preferences necessary.

Results of the stage III is a complete time table for the department (see Table 2).

Course	Teacher	Time	Room
4EK202C1	7	Tu 12:45	SB 327
4EK202C2	18	Tu 16:15	SB 107
4EK202C3	18	Tu 18:00	SB 206
4EK202C4	26	Mo 14:30	SB 206
4EK202C5	27	Th 12:45	SB 204
4EK211C1	11	Th 18:00	SB 108
4EK211C10	11	We 12:45	SB 212
4EK211C2	31	Th 16:15	SB 238

Course	Teacher	Time	Room
4EK213C2	27	Th 16:15	SB 207
4EK213C3	27	Th 14:30	SB 235
4EK213C4	26	Mo 11:00	NB C
4EK213P1	16	Fr 11:00	SB 206
4EK214C1	12	Mo 18:00	SB 206
4EK214C2	12	Mo 16:15	SB 108
4EK214P1	12	Mo 12:45	SB 409
4EK311C1	10	Tu 12:45	SB 206

Course	Teacher	Time	Room
4EK314C1	7	Tu 14:30	SB 237
4EK314P1	7	Tu 7:30	SB 238
4EK315C1	17	Tu 7:30	SB 108
4EK315C2	17	Tu 9:15	SB 206
4EK315P1	17	Mo 11:00	SB 227
4EK321C2	23	Th 11:00	CK 231
4EK411C1	10	Fr 11:00	SB 207
4EK411C2	19	Th 12:45	RB 114

Table 2 Sample of results of stage III

4 Conclusion

The strength of the sequential integer goal programming model is in combination of decomposed models with goal programming. The decomposition divides the scheduling problem into three simpler models. Goal programming enables using of soft constraints as goals instead of hard constraints usually used in timetabling models.

The time table obtained from the model modification ensures that all courses will be assigned by a teacher and a classroom. The model also avoids time conflicts. This means that the teacher will not teach two or more courses in the same time. Nevertheless, the schedule is not "good". In the model was not included preferences of teachers like they would like to teach e.g. only in three days of the week, they do not want to teach more than three courses in one day etc. The model also does not calculate with fact that some of the seminars should be thought at computer classrooms. If we add this condition to the stage III (seminars using computers have to be assigned to computer classrooms), the model does not have feasible solution. This condition has to be taken in to consideration in the previous stages. This might be the weakness of the Al-Husain's model, but it is solvable by another modification of the model. Other modifications might be added to solve the scheduling problem from the winter terms. In winter terms the classrooms assigned to the department take place in to different places in Prague, therefore we have to take into consideration the time for transfer between the places.

The future research will be focused on improving of this model to make "good schedules" for the department. Next step might be to help with the time table developing for the whole university. Another aim is to adapt the model for scheduling problem of secondary school.

Acknowledgements

The research is supported by the Internal Grant Agency of the University of Economics (grant no. F4/18/2011).

References

[1] Aladag, C., Hocaoglu, G., and Basaran, M.: The effect of neighborhood structures on tabu search algorithm in solving course timetabling problem. *Expert Systems with Applications* **36** (2009), 12349–12356.
[2] Al-Husain, R., Hasan, M. K., Al-Qaheri, H.: A Sequential Three-Stage Integer Goal Programming (IGP) Model for Faculty-Course-Time-Classroom Assignments. *Informatica* **35** (2011), 157-164.
[3] Bakir, M. A., Aksop, C.: A 0-1 integer programming approach to a university timetabling problem. *Hacettepe Journal of Mathematics and Statistics* **37** (2008), 41–55.
[4] Cerny, M.: *Computations*, Volume III (In Czech). Professional Publishing, Prague, 2012.
[5] Daskalaki, S., Birbas, T., Housos, E.: An integer programming formulation for a case study in university timetabling. *European Journal of Operational Research* **153** (2004), 117–135.
[6] De Causmaecker, P., Demeester, P., and Berghe, G.: A decomposed metaheuristic approach for a real-world university timetabling problem. *European Journal of Operational Research* **195** (2009), 307–318.
[7] Harwood, G. and Lawless, R.: Optimizing organizational goals in assigning faculty teaching schedules. *Decision Science* **6** (1975), 513–524.
[8] Van Dusseldorp, R. A. et al.: Applications of Goal Programming to Education. In: *Proceedings of the 14th Annual International Convention of the Association for Educational Data Systems* [online], Arizona, 1976. Available at http://www.eric.ed.gov/PDFS/ED127661.pdf [14th April 2011]

Risk-Sensitive and Average Optimality in Markov Decision Processes

Karel Sladký[1]

Abstract. This contribution is devoted to the risk-sensitive optimality criteria in finite state Markov Decision Processes. At first, we rederive necessary and sufficient conditions for average optimality of (classical) risk-neutral unichain models. This approach is then extended to the risk-sensitive case, i.e., when expectation of the stream of one-stage costs (or rewards) generated by a Markov chain is evaluated by an exponential utility function. We restrict ourselves on irreducible or unichain Markov models where risk-sensitive average optimality is independent of the starting state. As we show this problem is closely related to solution of (nonlinear) Poissonian equations and their connections with nonnegative matrices.

Keywords: dynamic programming, stochastic models, risk analysis and management.

JEL classification: C44, C61, C63
AMS classification: 90C40, 60J10, 93E20

1 Notation and Preliminaries

In this note, we consider unichain Markov decision processes with finite state space and compact action spaces where the stream of costs generated by the Markov processes is evaluated by an exponential utility function (so-called risk-sensitive models) with a given risk sensitivity coefficient.

To this end, let us consider an exponential utility function, say $\bar{u}^\gamma(\cdot)$, i.e. a separable utility function with constant risk sensitivity $\gamma \in \mathbb{R}$. For $\gamma > 0$ (risk averse case) $\bar{u}^\gamma(\cdot)$ is convex, if $\gamma < 0$ (risk seeking case) $\bar{u}^\gamma(\cdot)$ is concave. Finally if $\gamma = 0$ (risk neutral case) $\bar{u}^\gamma(\cdot)$ is linear. Observe that exponential utility function $\bar{u}^\gamma(\cdot)$ is separable and multiplicative if the risk sensitivity $\gamma \neq 0$ and additive for $\gamma = 0$. In particular, we have $u^\gamma(\xi_1 + \xi_2) = u^\gamma(\xi_1) \cdot u^\gamma(\xi_2)$ if $\gamma \neq 0$ and $u^\gamma(\xi_1 + \xi_2) \equiv \xi_1 + \xi_2$ for $\gamma = 0$.

Then the utility assigned to the (random) outcome ξ is given by

$$\bar{u}^\gamma(\xi) := \begin{cases} (\text{sign } \gamma)\, \exp(\gamma\xi), & \text{if } \gamma \neq 0, \\ \xi \quad \text{for } \gamma = 0. \end{cases} \tag{1}$$

For what follows let $u^\gamma(\xi) := \exp(\gamma\xi)$, hence $\bar{u}^\gamma(\xi) = (\text{sign } \gamma)\, u^\gamma(\xi)$. Obviously $\bar{u}^\gamma(\cdot)$ is continuous and strictly increasing. Then for the corresponding certainty equivalent, say $Z^\gamma(\xi)$, since $\bar{u}^\gamma(Z^\gamma(\xi)) = \mathsf{E}\,[\bar{u}^\gamma(\xi)]$ (E is reserved for expectation), we immediately get

$$Z^\gamma(\xi) = \begin{cases} \gamma^{-1} \ln\{\mathsf{E}\, u^\gamma(\xi)\}, & \text{if } \gamma \neq 0 \\ \mathsf{E}\,[\xi] & \text{for } \gamma = 0. \end{cases} \tag{2}$$

In what follows, we consider a Markov decision chain $X = \{X_n,\, n = 0, 1, \ldots\}$ with finite state space $\mathcal{I} = \{1, 2, \ldots, N\}$ and a compact set $\mathcal{A}_i$ 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 selected, then state j is reached in the next transition with a given probability $p_{ij}(a)$ and one-stage transition cost c_{ij} will be accrued to such transition.

[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: sladky@utia.cas.cz

A (Markovian) policy controlling the decision process is given by a sequence of decisions (actions) at every time point. In particular, policy controlling the process, $\pi = (f^0, f^1, \ldots)$, with $\pi^k = (f^k, f^{k+1}, \ldots)$ for $k = 1, 2, \ldots$, hence also $\pi = (f^0, f^1, \ldots, f^{k-1}, \pi^k)$ 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. 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 $p_{ij}(f_i)$.

Let $\xi_n^\alpha = \sum_{k=0}^{n-1} \alpha^k c_{X_k, X_{k+1}}$ with $\alpha \in (0,1)$, resp. $\xi_n = \sum_{k=0}^{n-1} c_{X_k, X_{k+1}}$, be the stream of α-discounted, resp. undiscounted, transition costs received in the n next transitions of the considered Markov chain X. Similarly let $\xi^{(m,n)}$ be reserved for the total (random) cost obtained from the mth up to the nth transition (obviously, $\xi_n = c_{X_0, X_1} + \xi^{(1,n)}$). Moreover, if the risk sensitivity $\gamma \neq 0$ then $\bar{u}^\gamma(\xi_n^\alpha) = (\text{sign } \gamma)\, u^\gamma(\xi_n^\alpha)$, resp. $\bar{u}^\gamma(\xi_n) = (\text{sign } \gamma)\, u^\gamma(\xi_n)$, is the (random) utility assigned to ξ_n^α, resp. to ξ_n. Observe that $\xi^\alpha := \lim_{n\to\infty} \xi_n^\alpha$ is well defined, hence $u^\gamma(\xi^\alpha) = \exp(\gamma \sum_{k=0}^{\infty} \alpha^k c_{X_k, X_{k+1}})$.

In the overwhelming literature on stochastic dynamic programming attention was mostly paid to the risk neutral case, i.e. if $\gamma = 0$. The following results and techniques adapted from [8] and [12] will be useful for derivation of necessary and sufficient average optimality conditions and for their further extensions to risk-sensitive models.

Introducing for arbitrary $g, w_j \in \mathbb{R}$ $(i, j \in \mathcal{I})$ the discrepancy function

$$\tilde{\varphi}_{i,j}(w, g) = c_{i,j} - w_i + w_j - g \tag{3}$$

we can easily verify the following identities:

$$\xi_n^\alpha = \frac{1-\alpha^n}{1-\alpha} g + w_{X_0} - \alpha^n w_{X_n} + \sum_{k=0}^{n-1} \alpha^k [\tilde{\varphi}_{X_k, X_{k+1}}(w, g) - (1-\alpha) w_{X_{k+1}}] \tag{4}$$

$$\xi_n = ng + w_{X_0} - w_{X_n} + \sum_{k=0}^{n-1} \tilde{\varphi}_{X_k, X_{k+1}}(w, g). \tag{5}$$

If the process starts in state i and policy $\pi = (f^n)$ is followed then for the expected α-discounted or undiscounted total cost $V_i^\pi(\alpha, n) := \mathsf{E}_i^\pi\, \xi_n^\alpha$, $V_i^\pi(n) := \mathsf{E}_i^\pi\, \xi_n$ we immediately get by (4)–(5)

$$V_i^\pi(\alpha, n) = \frac{1-\alpha^n}{1-\alpha} g + w_i + \mathsf{E}_i^\pi \{\sum_{k=0}^{n-1} \alpha^k [\tilde{\varphi}_{X_k, X_{k+1}}(w, g) - (1-\alpha) w_{X_{k+1}}] - \alpha^n w_{X_n}\} \tag{6}$$

$$V_i^\pi(n) = ng + w_i + \mathsf{E}_i^\pi \{\sum_{k=0}^{n-1} \tilde{\varphi}_{X_k, X_{k+1}}(w, g) - w_{X_n}\}. \tag{7}$$

Observe that

$$\mathsf{E}_i^\pi \sum_{k=0}^{n-1} \tilde{\varphi}_{X_k, X_{k+1}}(w, g) = \sum_{j \in \mathcal{I}} p_{ij}(f_i) \{\tilde{\varphi}_{i,j}(w, g) + \mathsf{E}_j^{\pi^1} \sum_{k=1}^{n-1} \tilde{\varphi}_{X_k, X_{k+1}}(w, g)\}. \tag{8}$$

It is well-known from the dynamic programming literature (cf. e.g. [1, 6, 9, 10, 16]) that

If there exists state $i_0 \in \mathcal{I}$ that is accessible from any state $i \in \mathcal{I}$ for every $f \in \mathcal{F}$ then $(*)$

(i) For every $f \in \mathcal{F}$ the resulting transition probability matrix $P(f)$ is *unichain* (i.e. $P(f)$ have no two disjoint closed sets),

(ii) There exists decision vector $\hat{f} \in \mathcal{F}$ along with numbers $\hat{w}_i, i \in \mathcal{I}$ (unique up to additive constant), and $\hat{g}$ being the solution of the set of (nonlinear) equations

$$\hat{w}_i + \hat{g} = \min_{a \in \mathcal{A}_i} \sum_{j \in \mathcal{I}} p_{ij}(a)[c_{i,j} + \hat{w}_j] = \sum_{j \in \mathcal{I}} p_{ij}(\hat{f}_i)[c_{i,j} + \hat{w}_j], \tag{9}$$

$$\varphi_i(f, \hat{f}) := \sum_{j \in \mathcal{I}} p_{ij}(f)[c_{i,j} + \hat{w}_j] - \hat{w}_i - \hat{g} \geq 0 \quad \text{with} \quad \varphi_i(\hat{f}, \hat{f}) = 0. \tag{10}$$

From (6), (7), (10) we immediately get for $V_i^\pi(\alpha) := \lim_{n\to\infty} V_i^\pi(\alpha, n)$

$$V_i^{\hat\pi}(\alpha) = \frac{1}{1-\alpha}\hat g + \hat w_i - (1-\alpha)\mathsf{E}_i^{\hat\pi}\sum_{k=0}^{\infty}\alpha^k \hat w_{X_{k+1}}, \quad V_i^{\hat\pi}(n) = n\hat g + \hat w_i - \mathsf{E}_i^{\hat\pi}\hat w_n$$

hence for stationary policy $\pi \sim (\hat f)$ and arbitrary policy $\pi = (f^n)$

$$\lim_{n\to\infty}\frac{1}{n}V_i^{\hat\pi}(n) = \lim_{\alpha\uparrow 1}(1-\alpha)V_i^{\hat\pi}(\alpha) = \hat g \tag{11}$$

$$\lim_{n\to\infty}\frac{1}{n}V_i^{\pi}(n) = \hat g = \lim_{\alpha\uparrow 1}(1-\alpha)V_i^{\pi}(\alpha) \quad \text{if and only if}$$

$$\lim_{n\to\infty}\frac{1}{n}\mathsf{E}_i^{\pi}\sum_{k=0}^{n-1}\varphi_{X_k}(f^n,\hat f) = 0. \tag{12}$$

2 Risk-Sensitive Optimality

On inserting the discrepancy function given by (3) in the exponential function $u^\gamma(\cdot)$ by (4) we get for the stream of discounted costs

$$\begin{aligned} u^\gamma(\xi_n^\alpha) &= \mathrm{e}^{\gamma\sum_{k=0}^{n-1}\alpha^k c_{X_k,X_{k+1}}} \\ &= \mathrm{e}^{\gamma[\sum_{k=0}^{n-1}\alpha^k g + w_{X_0} - \alpha^n w_{X_n}]} \times \mathrm{e}^{\gamma\sum_{k=0}^{n-1}\alpha^k[\tilde\varphi_{X_k,X_{k+1}}(w,g) - (1-\alpha)w_{X_{k+1}}]} \end{aligned} \tag{13}$$

and for $U_i^\pi(\gamma,\alpha,n) := \mathsf{E}_i^\pi u^\gamma(\xi_n^\alpha)$ we have

$$U_i^\pi(\gamma,\alpha,n) = \mathrm{e}^{\gamma[\frac{1-\alpha^n}{1-\alpha}g + w_i]} \times \mathsf{E}_i^\pi \mathrm{e}^{\gamma\{\sum_{k=0}^{n-1}\alpha^k[\tilde\varphi_{X_k,X_{k+1}}(w,g) - (1-\alpha)w_{X_{k+1}}] - \alpha^n w_{X_n}\}} \tag{14}$$

Observe that w_i's are bounded, i.e. $|w_{X_k}| \le K$ for some $K \ge 0$. Hence it holds

$$\mathrm{e}^{-|\gamma|K} \le \mathrm{e}^{\gamma(1-\alpha)\sum_{k=1}^{\infty}\alpha^k w_{X_{k+1}}} \le \mathrm{e}^{|\gamma|K} \tag{15}$$

and for n tending to infinity from (14) we immediately get for $U_i^\pi(\gamma,\alpha) := \lim_{n\to\infty} U_i^\pi(\gamma,\alpha,n)$

$$U_i^\pi(\gamma,\alpha) = \mathrm{e}^{\gamma[\frac{1}{1-\alpha}g + w_i]} \times \mathsf{E}_i^\pi \mathrm{e}^{\gamma\sum_{k=0}^{\infty}\alpha^k[\tilde\varphi_{X_k,X_{k+1}}(w,g) + (1-\alpha)w_{X_{k+1}}]} \tag{16}$$

Similarly for undiscounted models we get by (13), (14)

$$U_i^\pi(\gamma,n) = \mathrm{e}^{\gamma[ng + w_i]} \times \mathsf{E}_i^\pi \mathrm{e}^{\gamma[\sum_{k=0}^{n-1}\tilde\varphi_{X_k,X_{k+1}}(w,g) - w_{X_n}]} \tag{17}$$

Now observe that

$$\mathsf{E}_i^\pi \mathrm{e}^{\gamma\sum_{k=0}^{n-1}\tilde\varphi_{X_k,X_{k+1}}(w,g)} = \sum_{j\in\mathcal{I}} p_{ij}(f_i^0)\,\mathrm{e}^{\gamma[c_{i,j} - w_i + w_j - g]} \times \mathsf{E}_j^{\pi^1} \mathrm{e}^{\gamma\sum_{k=1}^{n-1}\tilde\varphi_{X_k,X_{k+1}}(w,g)} \tag{18}$$

$$\mathsf{E}_j^\pi\{\mathrm{e}^{\gamma\sum_{k=m}^{n-1}\tilde\varphi_{X_k,X_{k+1}}(w,g)} | X_m = j\} = \sum_{\ell\in\mathcal{I}} p_{j\ell}(f_j^m)\,\mathrm{e}^{\gamma[c_{j,\ell} - w_j + w_\ell - g]} \times \mathsf{E}_\ell^{\pi^{m+1}} \mathrm{e}^{\gamma\sum_{k=m+1}^{n-1}\tilde\varphi_{X_k,X_{k+1}}(w,g)} \tag{19}$$

Employing (18) the following facts can be easily verified by (16), (17).

Result 1.

(i) Let for a given stationary policy $\pi \sim (f)$ there exist $g(f)$, $w_j(f)$'s such that

$$\sum_{j\in\mathcal{I}} p_{ij}(f_i)\ \mathrm{e}^{\gamma[c_{i,j}-w_i(f)+w_j(f)-g(f)]} \ = \ \sum_{j\in\mathcal{I}} p_{ij}(f_i)\ \mathrm{e}^{\gamma\,\tilde{\varphi}_{i,j}(w(f),g(f))} = 1, \quad i\in\mathcal{I}. \tag{20}$$

Then

$$U_i^{\pi}(\gamma,\alpha) \ = \ \mathrm{e}^{\gamma(\frac{1}{1-\alpha}g(f)+w_i(f))} \times \mathsf{E}_i^{\,\pi}\,\mathrm{e}^{-\gamma\,(1-\alpha)\,\sum\limits_{k=1}^{\infty}\alpha^k\,w_{X_k}(f)} \tag{21}$$

$$U_i^{\pi}(\gamma,n) \ = \ \mathrm{e}^{\gamma(ng(f)+w_i(f))} \times \mathsf{E}_i^{\,\pi}\,\mathrm{e}^{-\gamma\ w_{X_n}(f)} \tag{22}$$

(ii) If it is possible to select $g=g^*$ and $w_j=w_j^*$'s, resp. $g=\hat{g}$ and $w_j=\hat{w}_j$'s, such that for any $f\in\mathcal{F}$, all $i\in\mathcal{I}$ and some $f^*\in\mathcal{F}$, resp. $\hat{f}\in\mathcal{F}$,

$$\sum_{j\in\mathcal{I}} p_{ij}(f_i)\ \mathrm{e}^{\gamma\,\tilde{\varphi}_{i,j}(w^*,g^*)} \le 1 \ \text{ with } \ \sum_{j\in\mathcal{I}} p_{ij}(f_i^*)\ \mathrm{e}^{\gamma\,\tilde{\varphi}_{i,j}(w^*,g^*)} = 1 \tag{23}$$

respectively

$$\sum_{j\in\mathcal{I}} p_{ij}(f_i)\ \mathrm{e}^{\gamma\,\tilde{\varphi}_{i,j}(\hat{w},\hat{g})} \ge 1 \ \text{ with } \ \sum_{j\in\mathcal{I}} p_{ij}(\hat{f}_i)\ \mathrm{e}^{\gamma\,\tilde{\varphi}_{i,j}(\hat{w},\hat{g})} = 1 \tag{24}$$

then

$$\mathrm{e}^{\gamma(\frac{1}{1-\alpha}\hat{g}+\hat{w}_i)}\cdot\mathrm{e}^{-|\gamma|K} \le U_i^{\hat{\pi}}(\gamma,\alpha) \le \ U_i^{\pi^*}(\gamma,\alpha) \le \ \mathrm{e}^{\gamma(\frac{1}{1-\alpha}g^*+w_i^*)}\cdot\mathrm{e}^{|\gamma|K} \tag{25}$$

$$\mathrm{e}^{\gamma(n\hat{g}+\hat{w}_i)}\cdot\mathrm{e}^{-|\gamma|K} \le U_i^{\hat{\pi}}(\gamma,n) \le \ U_i^{\pi^*}(\gamma,n)) \le \ \mathrm{e}^{\gamma(ng^*+w_i^*)}\cdot\mathrm{e}^{|\gamma|K}. \tag{26}$$

Result 2. Let (cf. (2)) $\quad Z_i^{\pi}(\gamma,\alpha)=\frac{1}{\gamma}\ln U_i^{\pi}(\gamma,\alpha), \quad Z_i^{\pi}(\gamma,n)=\frac{1}{\gamma}\ln U_i^{\pi}(\gamma,n).$

Then by (25), (26) for stationary policies $\hat{\pi}\sim(\hat{f})$, $\pi^*\sim(f^*)$, and by (16), (17) for an arbitrary policy $\pi=(f^n)$

$$\lim_{n\to\infty}\frac{1}{n}Z_i^{\hat{\pi}}(\gamma,n)=\lim_{\alpha\uparrow 1}(1-\alpha)Z_i^{\hat{\pi}}(\gamma,\alpha)=\hat{g}, \qquad \lim_{n\to\infty}\frac{1}{n}Z_i^{\pi^*}(\gamma,n)=\lim_{\alpha\uparrow 1}(1-\alpha)Z_i^{\pi^*}(\gamma,\alpha)=g^* \tag{27}$$

$$\lim_{n\to\infty}\frac{1}{n}Z_i^{\pi}(\gamma,n)=g^*, \quad \text{resp.} \quad \lim_{n\to\infty}\frac{1}{n}Z_i^{\pi}(\gamma,n)=\hat{g}, \quad \text{if and only if}$$

$$\lim_{n\to\infty}\frac{1}{n}\ln[\mathsf{E}_i^{\,\pi}\,\mathrm{e}^{\gamma\sum\limits_{k=0}^{n-1}\tilde{\varphi}_{X_k,X_{k+1}}(w^*,g^*)}]=0, \quad \text{resp.} \quad \lim_{n\to\infty}\frac{1}{n}\ln[\mathsf{E}_i^{\,\pi}\,\mathrm{e}^{\gamma\sum\limits_{k=0}^{n-1}\tilde{\varphi}_{X_k,X_{k+1}}(\hat{w},\hat{g})}]=0. \tag{28}$$

3 Poissonian Equations

The system of equations (20) for the considered stationary policy $\pi\sim(f)$ and the nonlinear systems of equations (23), (24) for finding stationary policy with maximal/minimal value of $g(f)$ can be also written as

$$\mathrm{e}^{\gamma[g(f)+w_i(f)]} \ = \ \sum_{j\in\mathcal{I}} p_{ij}(f_i)\,\mathrm{e}^{\gamma[c_{i,j}+w_j(f)]} \quad (i\in\mathcal{I}) \tag{29}$$

$$\mathrm{e}^{\gamma[g^*+w_i^*]} \ = \ \max_{f\in\mathcal{F}}\sum_{j\in\mathcal{I}} p_{ij}(f_i)\ \mathrm{e}^{\gamma[c_{i,j}+w_j^*]} \quad (i\in\mathcal{I}) \tag{30}$$

$$\mathrm{e}^{\gamma[\hat{g}+\hat{w}_i]} \ = \ \min_{f\in\mathcal{F}}\sum_{j\in\mathcal{I}} p_{ij}(f_i)\,\mathrm{e}^{\gamma[c_{i,j}+\hat{w}_j]} \quad (i\in\mathcal{I}) \tag{31}$$

respectively, for the values $g(f)$, $\hat{g}$, g^*, $w_i(f)$, w_i^*, $\hat{w}_i$ $(i=1,\dots,N)$; obviously, these values depend on the selected risk sensitivity γ. Eqs. (30), (31) can be called the *γ-average reward/cost optimality equation*. In particular, if $\gamma\downarrow 0$ using the Taylor expansion by (29), resp. (31), we have

$$g(f)+w_i(f)=\sum_{j\in\mathcal{I}} p_{ij}(f_i)\,[c_{i,j}+w_j(f)], \quad \text{resp.} \quad \hat{g}+\hat{w}_i=\min_{f\in\mathcal{F}}\sum_{j\in\mathcal{I}} p_{ij}(f_i)\,[c_{i,j}+\hat{w}_j]$$

that well corresponds to (9).

On introducing new variables $v_i(f) := \mathrm{e}^{\gamma w_i(f)}$, $\rho(f) := \mathrm{e}^{\gamma g(f)}$, and on replacing transition probabilities $p_{ij}(f_i)$'s by general nonnegative numbers defined by $q_{ij}(f_i) := p_{ij}(f_i) \cdot \mathrm{e}^{\gamma c_{ij}}$ (29) can be alternatively written as the following set of equations

$$\rho(f)v_i(f) \quad = \quad \sum_{j\in\mathcal{I}} q_{ij}(f_i)\, v_j(f) \quad (i \in \mathcal{I}) \tag{32}$$

and (30), (31) can be rewritten as the following sets of nonlinear equations (here $\hat{v}_i := \mathrm{e}^{\gamma \hat{w}_i}$, $v_i^* := \mathrm{e}^{\gamma w_i^*}$, $\hat{\rho} = \mathrm{e}^{\gamma \hat{g}}$, $\rho^* := \mathrm{e}^{\gamma g^*}$)

$$\rho^* v_i^* = \max_{f\in\mathcal{F}} \sum_{j\in\mathcal{I}} q_{ij}(f_i)\, v_j^*, \quad \hat{\rho}\,\hat{v}_i = \min_{f\in\mathcal{F}} \sum_{j\in\mathcal{I}} q_{ij}(f_i)\,\hat{v}_j \quad (i \in \mathcal{I}) \tag{33}$$

called *γ-average reward/cost optimality equation in multiplicative form.*

For what follows it is convenient to consider (32), (33) in matrix form. To this end, we introduce the $N \times N$ matrix $Q(f) = [q_{ij}(f_i)]$ with spectral radius (Perron eigenvalue) $\rho(f)$ along with its right Perron eigenvector $v(f) = [v_i(f_i)]$, hence (cf. [5]) $\rho(f)v(f) = Q(f)v(f)$. Similarly, for $v(f^*) = v^*$, $v(\hat{f}) = \hat{v}$ (33) can be written in matrix form as

$$\rho^* v^* = \max_{f\in\mathcal{F}} Q(f)v^*, \qquad \hat{\rho}\,\hat{v} = \min_{f\in\mathcal{F}} Q(f)\hat{v}. \tag{34}$$

Recall that vectorial maximum and minimum in (34) should be considered componentwise and $\hat{v}$, v^* are unique up to multiplicative constant.

Furthermore, if the transition probability matrix $P(f)$ is irreducible then also $Q(f)$ is irreducible and the right Perron eigenvector $v(f)$ can be selected strictly positive. To extend this assertion to unichain models in contrast to condition $(*)$ for the risk neutral case it is necessary to assume existence of state

$i_0 \in \mathcal{I}$ accessible from any state $i \in \mathcal{I}$ for every $f \in \mathcal{F}$ that belongs to the *basic* class[1] of $Q(f)$. $(**)$

If condition $(**)$ is fulfilled it can be shown (cf. [15]) that in (30), (31) and (33) eigenvectors $v(f)$, $\hat{v}$, v^* can be selected strictly positive and ρ^*, resp. $\hat{\rho}$, is the maximum, resp. minimum, Perron eigenvalue of the matrix family $\{Q(f), f \in \mathcal{F}\}$.

So we have arrived to

Result 3. Sufficient condition for the existence of γ-average reward/costs optimality equation is the existence of state $i_0 \in \mathcal{I}$ fulfilling condition $(**)$ (trivially fulfilled for irreducible models).

In particular, for unichain models condition $(**)$ is fulfilled if this risk sensitive coefficient γ is sufficiently close to zero (cf. [3, 4, 15]). Finding solution of (34) can be performed by policy or value iteration. Details can be found e.g. in [2, 3, 7, 14, 15].

Finally, we rewrite optimality condition (28) of Result 2 in terms of $Q(f)$, $\hat{v}$, $\hat{\rho}$. To this end first observe that

$$\begin{aligned} \mathsf{E}_{i_0}^{\pi}\, \mathrm{e}^{\gamma \sum_{k=0}^{n-1} \tilde{\varphi}_{X_k, X_{k+1}}(\hat{w},\hat{g})} &= \prod_{k=0}^{n-1} \left\{ \sum_{i_{k+1}\in\mathcal{I}} p_{i_k,i_{k+1}}(f_{i_k}^k)\, \mathrm{e}^{\gamma[c_{i_k,i_{k+1}} + \hat{w}_{i_{k+1}} - \hat{w}_{i_k} - \hat{g}]} \right\} \\ &= \prod_{k=0}^{n-1} \left\{ \sum_{i_{k+1}\in\mathcal{I}} q_{i_k,i_{k+1}}(f_{i_k}^k) \cdot \hat{v}_{i_{k+1}} \cdot \hat{v}_{i_k}^{-1} \cdot \hat{\rho}^{-1} \right\} \end{aligned} \tag{35}$$

So equation (28) can be also written in matrix form as

$$\lim_{n\to\infty} \frac{1}{n} \ln \left\{ \prod_{k=0}^{n-1} \hat{V}^{-1} Q(f^k) \hat{V} \hat{\rho}^{-1} \right\} = \lim_{n\to\infty} \frac{1}{n} \ln \left\{ \hat{V}^{-1} \left[\prod_{k=0}^{n-1} Q(f^k) \cdot \hat{\rho}^{-1} \right] \cdot \hat{V} \right\} \cdot \hat{V} = \mathbf{0} \tag{36}$$

where the diagonal matrix $\hat{V} = \mathrm{diag}\,[\hat{v}_i]$ and $\mathbf{0}$ is reserved for a null matrix.

[1] (i.e. irreducible class with spectral radius equal to the Perron eigenvalue of $Q(f)$)

4 Conclusions

In this note necessary and sufficient optimality conditions for discrete time Markov decision chains are obtained along with equations for average optimal policies both for risk-neutral and risk-sensitive models. Our analysis is restricted to unichain models, and for the risk-sensitive case some additional assumptions are made. For multichain models it is necessary to find suitable partition of the state space into nested classes that retain some properties of the unichain model. Some results in this direction can be found in [11, 15, 17, 18].

Acknowledgements

This research was supported by the Czech Science Foundation under Grants P402/11/0150 and P402/10/0956.

References

[1] Bertsekas, D. P.: *Dynamic Programming and Optimal Control. Volume 2. Third Edition.* Athena Scientific, Belmont, Mass. 2007.

[2] Cavazos-Cadena, R. and Montes-de-Oca R.: The value iteration algorithm in risk-sensitive average Markov decision chains with finite state space. *Mathematics of Operations Research* **28** (2003), 752–756.

[3] Cavazos-Cadena, R.: Solution to the risk-sensitive average cost optimality equation in a class of Markov decision processes with finite state space. *Mathematical Methods of Operations Research* **57** (2003), 253–285.

[4] Cavazos-Cadena, R. and Hernández-Hernández, D.: A characterization of the optimal risk-sensitive average cost infinite controlled Markov chains. *Annals of Applied Probability* **15** (2005), 175–212.

[5] Gantmakher, F. R.: *The Theory of Matrices.* Chelsea, London 1959.

[6] Howard, R. A.: *Dynamic Programming and Markov Processes.* MIT Press, Cambridge, Mass. 1960.

[7] Howard, R. A. and Matheson, J.: Risk-sensitive Markov decision processes. *Management Science* **23** (1972), 356–369.

[8] Mandl, P.: On the variance in controlled Markov chains. *Kybernetika* **7** (1971), 1–12.

[9] Puterman, M. L.: *Markov Decision Processes – Discrete Stochastic Dynamic Programming.* Wiley, New York 1994.

[10] Ross, S. M.: *Introduction to Stochastic Dynamic Programming.* Academic Press, New York 1983.

[11] Rothblum, U. G. and Whittle, P.: Growth optimality for branching Markov decision chains. *Mathematics of Operations Research* **7** (1982), 582–601.

[12] Sladký, K.: On the set of optimal controls for Markov chains with rewards. *Kybernetika* **10** (1974), 526–547.

[13] Sladký, K.: On dynamic programming recursions for multiplicative Markov decision chains. *Mathematical Programming Study* **6** (1976), 216–226.

[14] Sladký, K.: Bounds on discrete dynamic programming recursions I. *Kybernetika* **16** (1980), 526–547.

[15] Sladký, K.: Growth rates and average optimality in risk-sensitive Markov decision chains. *Kybernetika* **44** (2008), 205–226.

[16] Tijms, H. C.: *A First Course in Stochastic Models.* Wiley, Chichester, 2003.

[17] Whittle, P.: *Optimization Over Time – Dynamic Programming and Stochastic Control. Volume II, Chapter 35.* Wiley, Chichester 1983.

[18] Zijm, W. H. M.: *Nonnegative Matrices in Dynamic Programming.* Mathematical Centre Tract, Amsterdam 1983.

Modeling financial returns by discrete stable distributions

Lenka Slámová[1], Lev B. Klebanov[2]

Abstract. This paper develops a new approach to modeling financial returns by introducing discrete stable distributions. It is well known that the financial returns are not normally distributed, extremal events occur more often than the Gaussian distribution suggests. Already in the sixties Benoit Mandelbrot suggested a hypothesis that returns follow a stable Paretian law. Inspired by the discrete nature of prices appearing on the markets we model the financial returns by discrete analogues of absolutely continuous stable distributions. The known discrete stability of random variables on $\mathbb{N}$ is generalized to the case of random variables on $\mathbb{Z}$. We give brief introduction to the theory of discrete stability on $\mathbb{Z}$, show connection of discrete stable random variables to their absolutely continuous counterparts and focus mainly on methods of estimation of parameters of these distributions from the real data of financial returns.

Keywords: discrete stable distributions, parameter estimation, M-estimator

JEL classification: C13
AMS classification: 60E07

1 Introduction

The observation that financial returns have heavy tails and a hypothesis that they follow stable Paretian law were suggested by Benoît Mandelbrot. Since then many mathematicians and economists were studying the implications of this hypothesis, and stable distributions became a very important field of probability for both theoreticians and practitioners.

The stable Paretian hypothesis assumes that the financial returns are continuous random variables, which is the case of logarithmic returns ($\log(V_1/V_0)$) or arithmetic returns ($(V_1 - V_0)/V_0$). However sometimes it is more convenient to consider simple returns computed as a difference of buy price and sell price. These prices are quoted on the market on a discrete grid (sometimes called ticks). Then the returns are discrete random variable and thus the assumption they follow a stable Paretian law is incorrect. A discrete distribution that allows for heavy tails and have the stability property is needed for modeling such returns.

The notion of discrete stability for lattice random variables on positive integers was introduced in [5] and further studied in [1]. In [4] an extension of discrete stability for random variables on $\mathbb{Z}$ was proposed. A discrete analogue of stable Paretian distribution was introduced and a connection to absolutely continuous stable distributions was shown.

In this article we propose a method of parameter estimation for discrete stable distribution family and compare it with a well known empirical characteristic function method, that was reviewed in [6]. We illustrate the quality of both methods by estimating parameters of simulated data. Finally we do an empirical study on market data of futures prices and compare the performance of the fit with the normal and continuous stable distribution fit.

[1]Charles University in Prague, Department of probability and mathematical statistics, Sokolovská 83, 18675 Praha 8, e-mail: slamova.lenka@gmail.com
[2]Charles University in Prague, Department of probability and mathematical statistics, Sokolovská 83, 18675 Praha 8, e-mail: klebanov@chello.cz

2 Discrete stable distributions

A random variable X is strictly stable if for all $n \in \mathbb{N}$ there is a constant a_n such that $a_n \sum_{i=1}^{n} X_i \stackrel{d}{=} X$, where X_i's are independent copies of X. When defining stability for discrete random variables one has to choose a different normalization, because the normalized sum $a_n \sum_{i=1}^{n} X_i$ is not generally integer valued.

A definition of positive discrete stable random variables was introduced in [5], where a normalization $\sum_{i=1}^{n} \tilde{X}_i(p_n) \stackrel{d}{=} X$ was used, with $\tilde{X}(p_n) = \sum_{j=1}^{X} \varepsilon_j$, where $\mathbb{P}(\varepsilon_i = 1) = 1 - \mathbb{P}(\varepsilon_i = 0) = p_n$ with $p_n \in (0, 1)$. A positive random variable X with characteristic function

$$f(t) = \exp\left\{-\lambda(1 - e^{\mathrm{i}t})^{\gamma}\right\}, \quad \lambda > 0, \gamma \in (0, 1]$$

is called *positive discrete stable* and is denoted by $\mathscr{PDS}(\gamma, \lambda)$. This distribution has its support in $\mathbb{N}_0$ and in case of $\gamma = 1$ corresponds to Poisson distribution. [1] showed that positive discrete stable distribution with $\gamma < 1$ belongs to the domain of normal attraction of the absolutely continuous positive stable distribution (stable distribution with index of stability $\alpha = \gamma$ and with skewness parameter $\beta = 1$), whose Laplace transform is given by $\exp(-\lambda t^{\gamma})$. The case of $\gamma = 1$ is a degenerate one, where the normalized sum of Poisson random variables converges to a constant λ.

The discrete stability for random variables on $\mathbb{Z}$ was defined in [4]. The first approach to symmetric discrete random variables uses the following normalization:

$$\sum_{j=1}^{n} \tilde{X}_j(p_n) \xrightarrow{d} X, \quad \text{where} \quad \tilde{X}(p_n) = \sum_{i=1}^{|X|} \varepsilon_i(p_n)$$

and $\varepsilon_i(p_n)$ are taking values ± 1 with probability p_n and 0 with probability $1 - 2p_n$, with $p_n \downarrow 0$. This normalization leads to a distribution with characteristic function

$$f(t) = \exp\left\{-\lambda\left(1 - \frac{1}{2}\left(e^{\mathrm{i}t} + e^{-\mathrm{i}t}\right)\right)^{\gamma}\right\} \quad \text{with} \quad \lambda > 0, \gamma \in (0, 1].$$

Such a distribution is then called *symmetric discrete stable* and is denoted by $\mathscr{SDS}(\gamma, \lambda)$. This distribution belongs to the domain of normal attraction of the absolutely continuous symmetric stable distribution with index of stability $\alpha = 2\gamma$. Hence the case $\gamma = 1$ can be considered as a discrete version of Gaussian distribution.

The general case of discrete random variables on $\mathbb{Z}$ requires once again a different normalization, namely

$$\sum_{j=1}^{n} \tilde{X}_j(p_n^1, p_n^2) \xrightarrow{d} X, \quad \text{where} \quad \tilde{X}(p_n^1, p_n^2) = \begin{cases} \sum_{i=1}^{X} \varepsilon_i(p_n^1), & X \geq 0 \\ -\sum_{i=1}^{|X|} \varepsilon_i(p_n^2), & X < 0 \end{cases}$$

and $\varepsilon_i(p)$ are taking values 1 with probability p and 0 with probability $1 - p$, where $p_n^i \downarrow 0$. The characteristic function of such distribution takes the following form

$$f(t) = \exp\left\{-\lambda_1\left(1 - e^{\mathrm{i}t}\right)^{\gamma} - \lambda_2\left(1 - e^{-\mathrm{i}t}\right)^{\gamma}\right\}, \quad \text{with} \quad \lambda_1, \lambda_2 > 0, \gamma \in (0, 1].$$

Such a distribution is called *discrete stable distribution* and is denoted by $\mathscr{DS}(\gamma, \lambda_1, \lambda_2)$. The discrete stable distribution for $\gamma < 1$ is in the domain of normal attraction of absolutely continuous stable distribution with index of stability $\alpha = \gamma$ and for $\gamma = 1$ in the domain of normal attraction of symmetric stable distribution with index of stability $\alpha = 2$.

3 Parameter estimation

Statistical methods of estimation in case of discrete stable distributions has several problems that inhibit to use most of the methods - like maximum likelihood or method of moments. These methods are based on assumptions like availability of a closed form of probability function or existence of moments up to some order. However this is not the case of discrete stable distributions. We describe here and use ECF method using empirical characteristic function and AML method which use M-estimator for estimation of the parameters of discrete stable distributions.

3.1 Empirical characteristic function method

This method was studied broadly as an alternative method of estimation when the cumulative distribution function is not known in a closed form, but we have a closed form for characteristic function. [3] introduced an ECF method known as "$k-L$ procedure". A review of different approaches to ECF method can be found in [6].

Let $\{P_\theta, \theta \in \Theta\}$ be a family of probability distributions on a probability space $(\mathcal{X}, \mathcal{M})$, where the parametric space $\Theta = \mathbb{R}^d$. The characteristic function of the distribution P_θ is defined by $f(t,\theta) = \mathbb{E}_\theta\left[e^{\mathrm{i}tX}\right]$ and the empirical characteristic function (ECF) based on a sample of observations $x_1, \ldots, x_n$ is defined by

$$\widehat{f}_n(t) = \frac{1}{n}\sum_{i=1}^{n} e^{\mathrm{i}tx_i}.$$

The general idea for ECF estimation is to minimize distance between f and $\widehat{f}_n$. Since discrete stable distributions are lattice distributions, their characteristic function is periodic with period 2π, it is enough to to minimize the distance only on the interval $[-\pi, \pi]$. Choose a discrete grid of points $\{t_1, \ldots, t_k\} \in [-\pi, \pi]$ and denote by

$$V_n = \big(\widehat{f}_n(t_1), \ldots, \widehat{f}_n(t_k)\big)' \quad \text{and} \quad V_\theta = \big(f(t_1), \ldots, f(t_k)\big)'.$$

The estimate that solves $\min_\theta (V_n - V_\theta)'(V_n - V_\theta)$ corresponds to the nonlinear OLS regression of V_n of V_θ. [6] argues that this estimator is not efficient and [3] suggests to use nonlinear GLS regression as follows. Denote by Ω the covariance matrix of V_n, then an efficient estimator can be obtained as a solution of

$$\min_\theta (V_n - V_\theta)'\widehat{\Omega}^{-1}(V_n - V_\theta),$$

where $\widehat{\Omega}$ is a consistent estimate of Ω.

3.2 Approximate maximum likelihood method

Let $\{P_\theta, \theta \in \Theta\}$ be a family of probability distributions on a probability space $(\mathcal{X}, \mathcal{M})$, where the parametric space $\Theta = \mathbb{R}^d$. In the maximum likelihood estimation one assumes the existence of a density $p(x,\theta)$ and of a function

$$\mathbf{J}(x,\theta) = \left(\frac{\frac{\partial p}{\partial \theta_i}(x,\theta)}{p(x,\theta)}\right)_{i=1,\cdots,d}.$$

The maximum likelihood estimator θ^* of the parameter θ, given a set of n observations $x_1, \cdots, x_n$, is a solution of $\sum_{j=1}^n \mathbf{J}(x_j,\theta) = \mathbf{0}$. However if the density does not exist or it is not known in an analytical form, this method cannot be used. Let us denote by $\mathcal{L}_k$ a linear space generated by set of functions $\{1, \varphi_1(x), \cdots, \varphi_k(x)\}$ on $(\mathcal{X}, \mathcal{M})$. We assume we know the functionals of our distribution, namely

$$\begin{aligned} \mathbb{E}_\theta \varphi_i(X) &= \pi_i(\theta), \quad i = 0, \ldots, k, \\ \mathbb{E}_\theta \varphi_i(X)\varphi_j(X) &= \pi_{ij}(\theta), \quad i,j = 0, \ldots, k. \end{aligned}$$

Instead of the function $\mathbf{J}$ we define a new function $\widehat{\mathbf{J}}(x,\theta)$ as a projection of $\mathbf{J}$ onto the space $\mathcal{L}_k$. Then $\widehat{\mathbf{J}}$ has to take the following form

$$\widehat{\mathbf{J}}(x,\theta) = \mathbf{c}(\theta)\varphi(x),$$

where $\varphi(x) = \big(\varphi_j(x), j = 0, \ldots, k\big)$, $\mathbf{c}(\theta) = \big(c_{ij}(\theta), i = 1, \ldots, d, j = 0, \cdots, k\big)$.

Since $\widehat{\mathbf{J}}$ is a projection of $\mathbf{J}$ onto $\mathcal{L}_k$, it has to hold

$$\mathbb{E}_\theta\left(J_i(X,\theta) - \widehat{J}_i(X,\theta)\right)\varphi_j(x) = 0, \quad i = 1, \ldots, d; j = 0, \ldots, k.$$

From this it follows that it has to hold

$$\frac{\partial \pi_j}{\partial \theta_i}(\theta) = \sum_{m=0}^{k} c_{im}(\theta)\pi_{mj}(\theta), \quad i = 1, \ldots, d; j = 0, \ldots, k.$$

If we use a matrix notation, $\nabla\pi(\theta) = \mathbf{c}(\theta)\Pi(\theta)$. Hence if the inverse of $\Pi(\theta)$ exists, then we are able to compute the matrix $\mathbf{c}(\theta) = \nabla\pi(\theta)\left(\Pi(\theta)\right)^{-1}$. Now the M-estimator θ^* of the parameter θ, given n observations $x_1, \ldots, x_n$, is the solution of

$$\sum_{k=1}^{n} \widehat{\mathbf{J}}(x_k, \theta) = \mathbf{0} \quad \text{or} \quad \sum_{m=1}^{n}\sum_{j=0}^{k} c_{ij}(\theta)\varphi_j(x_m) = 0, \quad \text{for} \quad i = 1, \ldots, d.$$

The question that arises with this method is the choice of functions φ. One possibility is as follows. Choose $k \in \mathbb{N}$ and then $z_1, \ldots, z_k \in \mathbb{Z}$ and define $\varphi_i(x) = z_i^x$ for $i = 1, \ldots, k$. Then the functionals $\pi_i(\theta) = \mathbb{E}_\theta(z_i^X) = \mathcal{P}(z_i)$, where $\mathcal{P}$ is the probability generating function of our distribution. A wise choice of k and z_i's is such that the variance of the resulting estimator is minimal.

4 Simulation from discrete stable family

The simulation from positive discrete stable distribution uses the stochastic representation emphasized by [2] stating that

$$\mathscr{PDS}(\gamma, \lambda) \stackrel{d}{=} \mathscr{P}\big(\lambda^{1/\gamma}\mathscr{S}(\gamma, 1, \sigma, 0)\big),$$

where $\mathscr{P}(\lambda)$ is Poisson r.v. with parameter λ, $\mathscr{S}(\alpha, \beta, \sigma, \mu)$ is stable r.v. with index of stability α, skewness β, scale σ and location μ, and $\sigma = (\cos(\gamma\pi/2))^{1/\gamma}$.

For simulation from discrete stable distribution we use the fact that

$$\mathscr{DS}(\gamma, \lambda_1, \lambda_2) \stackrel{d}{=} \mathscr{PDS}(\gamma, \lambda_1) - \mathscr{PDS}(\gamma, \lambda_2).$$

Symmetric discrete stable distribution have a similar stochastic representation as the positive case.

Theorem 1. *A symmetric discrete stable random variable $\mathscr{SDS}(\gamma, \lambda)$ is distributed as a compound Poisson random variable Y with intensity $\lambda^{1/\gamma}\mathscr{S}(\gamma, 1, \sigma, 0)$, jumps taking values ± 1 with the same probability and $\sigma = (\cos(\gamma\pi/2))^{1/\gamma}$.*

Proof. The Laplace transform of a random variable $\mathscr{S}(\gamma, 1, \sigma, 0)$ is $\mathbb{E}\exp(-u\mathscr{S}(\gamma, 1, \sigma, 0)) = \exp(-u^\gamma)$. The characteristic function of a compound Poisson variable with intensity λ and jumps taking values ± 1 with the same probability is $\exp(-\lambda(1 - 1/2(e^{\mathrm{i}t} + e^{-\mathrm{i}t})))$. Then it follows easily that the characteristic function of the random variable Y is

$$\mathbb{E}\left[e^{\mathrm{i}tY}\right] = \mathbb{E}\exp\left\{-\lambda^{1/\gamma}\mathscr{S}(\gamma, 1, \sigma, 0)\left(1 - \frac{1}{2}\left(e^{\mathrm{i}t} + e^{-\mathrm{i}t}\right)\right)\right\} = \exp\left\{-\lambda\left(1 - \frac{1}{2}\left(e^{\mathrm{i}t} + e^{-\mathrm{i}t}\right)\right)^\gamma\right\}.$$

And in this we recognize characteristic function of $\mathscr{SDS}(\gamma, \lambda)$. □

This representation gives us a useful tool for simulation from symmetric discrete stable distribution since stable and compound Poisson random variable may be easily generated.

5 Simulation study

In this section we will compare the performance of the two methods, ECF and AML, on simulated data. We simulate samples from positive discrete stable, symmetric discrete stable and discrete stable distributions with different values of parameters. The results are in Table 1. We can see that the AML method is better for estimating the index of stability γ, however the performance of the ECF method is better with the scale parameter λ. This is evident especially in the $\mathscr{DS}$ case, where the ECF method gives a very biased estimate of γ.

$\mathscr{PDS}$	(γ, λ)	(1, 0.5)	(0.8, 2)	(0.5, 4)	(0.2, 1)
Method	ECF	(1.0044, 0.5011)	(0.8052, 2.0220)	(0.4997, 3.9903)	(0.2052, 1.0098)
	AML	(1.0000, 0.4999	(0.8043, 2.0242)	(0.4998, 3.9725)	(0.2050, 1.0093)
$\mathscr{SDS}$	(γ, λ)	(1, 0.5)	(0.8, 2)	(0.5, 4)	(0.2, 1)
Method	ECF	(1.016, 0.4995)	(0.7886, 2.0389)	(0.4968, 4.0751)	(0.2041, 0.9960)
	AML	(1.000, 0.4972)	(0.8054, 2.0810)	(0.4977, 4.0823)	(0.1997, 0.9887)

$\mathscr{DS}$	$(\gamma, \lambda_1, \lambda_2)$	(1, 1, 2)	(0.5, 3, 0.5)	(0.3, 1, 1)
Method	ECF	(0.499, 0.981, 1.965)	(0.246, 2.934, 0.467)	(0.151, 0.987, 1.027)
	AML	(1.000, 0.856, 1.876)	(0.492, 2.935, 0.477)	(0.299, 1.000, 1.004)

Table 1: Estimated parameters of $\mathscr{PDS}$, $\mathscr{SDS}$ and $\mathscr{DS}$ distribution from simulated data

6 Empirical application to market data

As we mentioned in the introduction, one of the motivation for studying discrete stable distributions is an effort to have a more appropriate tool for modeling discrete financial returns. In this section we take a look at real data and show that discrete stable distributions are able to capture the nature of the data really well. We work with Bund futures prices (i.e. futures on German government bonds with maturity 8.5 to 10.5 years) from May 2010 and we consider intraday returns in ticks over different time periods (30 seconds, 1 minute, 2 minutes, 5 minutes, 15 minutes, 30 minutes, 1 hour). Such short returns are of interest for example for market makers and high frequency trading systems. The estimated parameters of the $\mathscr{SDS}$ and $\mathscr{DS}$ distribution by the AML method are displayed in Table 2. It is interesting to notice that the returns over different periods keep the same index of stability and the scale parameter changes, what suggest that the data have the stability property.

$\mathscr{SDS}$	period	10 s	30 s	1 min	2 min
Method	AML	(0.888, 0.760)	(0.852, 1.552)	(0.834, 2.495)	(0.854, 4.551)
$\mathscr{SDS}$	period	5 min	15 min	30 min	1 hour
Method	AML	(0.836, 8.863)	(0.796, 16.303)	(0.779, 25.946)	(0.764, 40.083)

$\mathscr{DS}$	period	10 s	30 s	1 min	2 min
Method	AML	(0.989, 0.913, 0.656)	(0.955, 0.816, 0.808)	(0.959, 1.414, 1.393)	(0.968, 2.577, 2.515)
$\mathscr{DS}$	period	5 min	15 min	30 min	1 hour
Method	AML	(0.928, 3.445, 3.401)	(0.864, 4.112, 4.057)	(0.816, 4.695, 4.414)	(0.777, 5.199, 4.679)

Table 2: Estimated parameters of $\mathscr{SDS}(\gamma, \lambda)$ and $\mathscr{DS}(\gamma, \lambda_1, \lambda_2)$ from real data

The quality of the fit of the empirical data with symmetric discrete stable, discrete stable and stable distribution is displayed at Figure 1. The mean square error of the fit with $\mathscr{SDS}$ distribution is 1.4e-5, for $\mathscr{DS}$ is 1.6e-5 and with $\mathscr{S}$ is 3.0e-5.

Acknowledgements

Authors were partially supported by Czech Science Foundation (grant P402/12/G097 and grant P203/12/0665).

References

[1] Christoph, G., and Schreiber, K.: Discrete stable random variables. *Statistics & Probability Letters* **37** (1998), 243–247.

[2] Devroye, L.: A triptych of discrete distributions connected to stable law. *Statistics & Probability Letters* **18** (1993), 349–351.

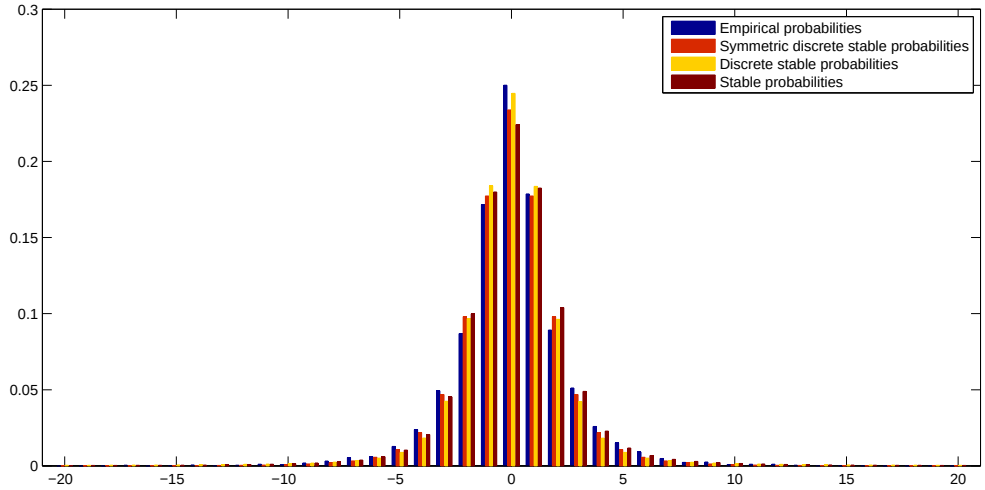


Figure 1: Empirical probabilities together with the theoretical probabilities of the fitted distributions

[3] Feuerverger, A., and McDunnough, P.: On the efficiency of empirical characteristic function procedure. *Journal of the Royal Statistical Society B* **43** (1981), 20–27.

[4] Klebanov, L., and Slámová, L.: Integer valued stable random variables. *Submitted for publication* (2012).

[5] Steutel, F. W., and van Harn, K. : Discrete analogues of self-decomposability and stability. *Annals of Probabability* **7** (1979), 893–899.

[6] Yu, J.: Empirical characteristic function estimation and its applications. *Econometric reviews* **23** (2004), 93–123.

A Proposal of Flexible Trend Specification in DSGE Models

Martin Slanicay[1]

Abstract. In this paper I propose flexible trend specification for estimating DSGE models on the log differences. I demonstrate this flexible trend specification on a New Keynesian DSGE model of two economies, which I consequently estimate on the data of the Czech economy and Euro Area 12, using Bayesian techniques. The advantage of the proposed trend specification is that the trend component and the cyclical component are modeled jointly in one model and the model itself decides which part of the data belongs to the trend component and which part belongs to the cyclical component. The proposed trend specification is flexible in the sense that smoothness of the trend can be easily modified by different calibration of the trend parameters. Results suggest that this method is able to find very reasonable trend in the data. Moreover, according to the Bayes factor the proposed specification decisively outperforms the original specification of the model estimated on the demeaned log differences with the same number of shocks.

Keywords: DSGE model, trend specification, Bayesian estimation.

JEL classification: C51, C68, E32
AMS classification: 91B51, 91B64

1 Motivation

DSGE models are models of cyclical fluctuations of the economy, therefore most of them can be estimated only on the stationary data. However, most economic time series are non-stationary, contain trends or display breaks. It implies that a transformation of the data is necessary in order to make them stationary. There are many different approaches which try to extract the cyclical components from the data and thus make the data stationary. However, each method extracts a different type of information from the data and there is no professional consensus on of what constitutes business cycle fluctuations. As a result, the stylized facts about the business cycle seems to differ substantially among detrending methods, even qualitatively, see Canova [7]. It seems that there is no general rule governing the transformation of time series and the suitability of each method depends on the particular situation and the research purpose.

The goal of this paper is to propose a flexible trend specification for estimating DSGE models on log differences. I demonstrate this flexible trend specification on New Keynesian DSGE model of two economies, presented in Kolasa [11], which I estimate on the data of the Czech economy and Euro Area 12, using Bayesian techniques.

2 Literature Review

Basically, there are three possible approaches for decomposition of time series into the trend and cyclical components, (i) detrend the actual data, (ii) build-in a trend into the model and (iii) use data transformations which, in theory, are likely to be void of non-cyclical fluctuations.

First approach implies that the data are detrended out of the model and the model is then estimated on these detrended data. Mostly, these filtered trends are free of any economic interpretation. Advantage of these methods is their universality and, in some cases (e.g. HP filter), simple implementation. On the other hand, one can argue that an arbitrary choice of a detrending method can significantly change the

[1]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, Brno 602 00, slanicay@mail.muni.cz

behavior of the model and obtained results can substantially differ among different detrending methods, see Canova [7], [8]. Moreover, many detrending methods are also applied on each time series individually and these filtered trends can be inconsistent with each other. Canova and Ferroni [9] propose a new method for estimating DSGE models based on combining the information provided by a variety of filters. They consider data filtered with alternative procedures as contaminated proxies of the relevant model-based variables and estimate structural and nonstructural parameters jointly using a signal extraction approach.

Second approach means that the data are detrended within the model. Advantage of this approach is that the decomposition between the trend and cyclical components of the data is performed by the model itself. It means, that the model itself decides which part of the data belongs to the trend component and which part belongs to the cyclical component. Nevertheless, there are still many issues, which depend on the researcher's choice, such as some assumptions about the trend specification or the relationship between the trend and cyclical component, which makes the previous objection concerning arbitrary choice of a detrending method still, however less, valid. Aguiar and Gopinath [2] show that shocks to trend growth, rather than transitory fluctuations around a stable trend, are the primary source of fluctuations in emerging economies. It can be seen as an argument for explicit modeling of the trend component within the model. Andrle [3] also argues for incorporating explicit (possibly structural) assumptions on trend behavior. He argues that permanent shocks influence business cycle behavior and ad-hoc detrended models *must have hard times to explain the comovement of the data.*[1] Brůha [4] proposes a small labor market model where he jointly models the trends and the cycles in a way that is slightly similar to the approach proposed in this paper. Canova [8] proposes a new method for estimating cyclical DSGE models using the raw data. This method is based on the flexible specification of the trend, which does not require that the cyclical component is solely located at business cycle frequencies.

Third approach is based on the fact that some transformations of the data may display fluctuations around some stable value and after removing this stable value, which can be regarded as a steady state, the data may look like stationary. Very popular is a transformation of the data using log differences, where steady state values of these log differences can be interpreted as a steady state growth rate. For example see Smets and Wouters [14] where is presented the policy model of ECB or Adolfson et. al. [1] where is presented the policy model "RAMSES" of the Swedish central bank. Another popular method can be found in Cogley [10] and McGrattan [13]. They suggest to estimate the model using the data in the form of real "great ratios", i.e. shares of real consumption (investment, etc.) on the real GDP, which exploits the fact that (in some countries) these shares are very stable. Therefore, after removing steady state values of these shares, the resulting deviations should look like stationary. Similarly, Whelan [15] suggests to estimate the model using the data in the form of nominal "great ratios", i.e. using the share of nominal consumption (investment, etc.) on the nominal GDP. The main pitfall of this method is that in many countries these ratios are not so stable, therefore the resulting deviations of these ratios from their "steady steady" values do not resemble the stationary data at all.

3 Model

I use a New Keynesian DSGE model of two economies, presented in Kolasa [11]. Derivation of this model from microfoundations, as well as its log-linear form can be found in Kolasa [11]. In this section I restrict my description of this model to a brief non-technical overview of its structure.

The model is a two-country model where both economies are modeled in the same way. The problematic fact that one economy is much more smaller than the other one is solved by the parameter n which governs the relative size of both economies.

The model assumes 5 types of representative agents in both economies. 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, following Burstein et al. [5]. Households also trade bonds and their intertemporal choice about consumption is influenced by preference shocks. Households supply labor and set wages on the monopolistically-competitive 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 Calvo constraint on the frequency of wage adjustment, see Calvo [6]. According to the Calvo constraint, every

[1] Andrle [3], p. 1

period each household resets its wage with the probability $1-\theta_w$ and with the probability θ_w partially adjust its wage according past inflation. Households also accumulate capital which they rent to firms. Capital accumulation is subject to investment-specific technological shocks and adjustment costs.

There are two types of firms in the model, 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 the monopolistically-competitive good markets. They set prices on the good market subject to a set of demand constraints and to the Calvo constraint on the frequency of price adjustment, see Calvo [6]. According to the Calvo constraint, every period each firm resets its price with the probability $1-\theta_p$ and with the probability θ_p partially adjust its price according past inflation.

Fiscal authority collects lump-sum taxes which they use for government expenditures and transfers to households, so that the state budget is balanced each period. The government expenditures consist only of domestic non-tradable goods and are modeled as a stochastic process - government expenditures shock. Given our assumptions about households, Ricardian equivalence holds in this model. Monetary authority behaves according a backward-looking Taylor rule and deviations from this rule are explained as monetary shocks. The model is completed with an assumption of a complete bond market and an assumption of goods and labor markets clearing.

The behavior of the model is driven by seven structural shocks in both economies: productivity shock in tradable sector and non-tradable sector, labor supply shock, investment efficiency shock, consumption preference shock, government spending shock and monetary policy shock. Except for the monetary policy shock, which is modeled as IID process, all shocks are represented by AR1 process. I allow for correlations between corresponding shocks in both economies, e.g. between domestic preference shocks and foreign preference shocks.[2]

4 Data and Trends

The model is estimated using quarterly data of the Czech economy and Euro Area 12 economy from the 1.Q 2000 to 3.Q 2011. Data series are downloaded from the web database of Eurostat. I use following 14 time series (seven for each economy): real GDP (y), consumption (c), investment (i), HICP (p), real wage (w), short-term interest rate (r) and internal exchange rate (x) defined as prices of non-tradable goods (services and energy) relative to prices of tradable goods (others). Except for the nominal interest rates, all observables are seasonally adjusted and expressed as log differences. Nominal interest rate is expressed as quarterly rate.

Let's now describe the decomposition of the observables into the trend and cyclical components. u_t^{obs} and r_t^{obs} denote the observables, where $u \in \{y, c, i, p, w, x\}$; γ_t^u and γ_t^r denote the trend components; u_t and r_t denote the cyclical components; and ε_t^u and ε_t^r are the trend shocks. Consequently, the decomposition of the observables can be written in the following form

$$
\begin{aligned}
u_t^{obs} &= \gamma_t^u + u_t - u_{t-1}, & r_t^{obs} &= \gamma_t^r + r_t,\\
\gamma_t^u - \gamma^u &= \rho_u(\gamma_{t-1}^u - \gamma^u) + \varepsilon_t^u, & \gamma_t^r - \gamma^r &= \rho_r(\gamma_{t-1}^r - \gamma^r) + \varepsilon_t^r,\\
\gamma^u &= mean(u_t^{obs}), \quad \rho_u \in (0,1), & \gamma^r &= mean(r_t^{obs}), \quad \rho_r \in (0,1).
\end{aligned}
$$

The interpretation of the proposed trend specification is straightforward. If the observable u_t^{obs} (expressed as growth rate) displays significant deviation from its steady state (given by the average growth rate γ^u), than the model explains this deviation partly by the temporary trend shock ε_t^u and partly by the cyclical component u_t. However, effect of this temporary trend shock is, to a certain degree, persistent, which means that the effect of such shock does not disappear immediately, but it takes some time. The degree of persistence of such shock is given by the parameters ρ_u. Note that while these trend shocks ε_t^u have temporary effects on the growth rates, they have permanent effects on the levels. As regards decomposition of the interest rate r_t^{obs}, the interpretation is very similar. If the interest rate r_t^{obs} displays significant deviation from its steady state (given by the average interest rate γ^r), than the model explains this deviation partly by the temporary trend shock ε_t^r and partly by the cyclical component r_t. However, effect of this temporary trend shock is, to a certain degree, persistent, which means that the

[2] For shocks represented by AR1 process it means that I allow for correlations between the innovations in these shocks.

effect of such shock does not disappear immediately, but it takes some time. The degree of persistence of such shock is given by the parameters ρ_r.

The advantage of the proposed trend specification is that the trend component and the cyclical component are modeled jointly in one model and the model itself decides which part of the data belongs to the trend component and which part belongs to the cyclical component. Nevertheless, there are still several things which depend on the researcher's choice. It is obvious that the standard deviations of trend shocks ε_t^u and ε_t^r and the persistence parameters ρ_u and ρ_r can not be successfully estimated together, because of the lack of identifiability. Therefore, I decided to calibrate the standard deviations of trend shocks ε_t^u and ε_t^r to one third of the standard deviations of the observables u_t^{obs} and r_t^{obs}; and to estimate persistence parameters ρ_u and ρ_r with prior distribution Beta, prior mean equal to 0.7, and prior std. deviation equal to 0.1. That is why I label the proposed trend specification as "flexible", because smoothness of the trend can be easily modified by the different calibration of the standard deviations of trend shocks ε_t^u and ε_t^r and by the prior setting of the persistence parameters ρ_u and ρ_r.

5 Estimation Results

The model is estimated with Random Walk Chain Metropolis-Hastings algorithm, using Dynare toolbox for Matlab. I generated two 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 25%, which is in line with the informal recommendation about ideal acceptance rate, see for example Koop [12]. According to the MCMC convergence diagnostics, checkplots, smoothed shocks and variables, and the prior and posterior distributions of the parameters, the model was estimated successfully.[3]

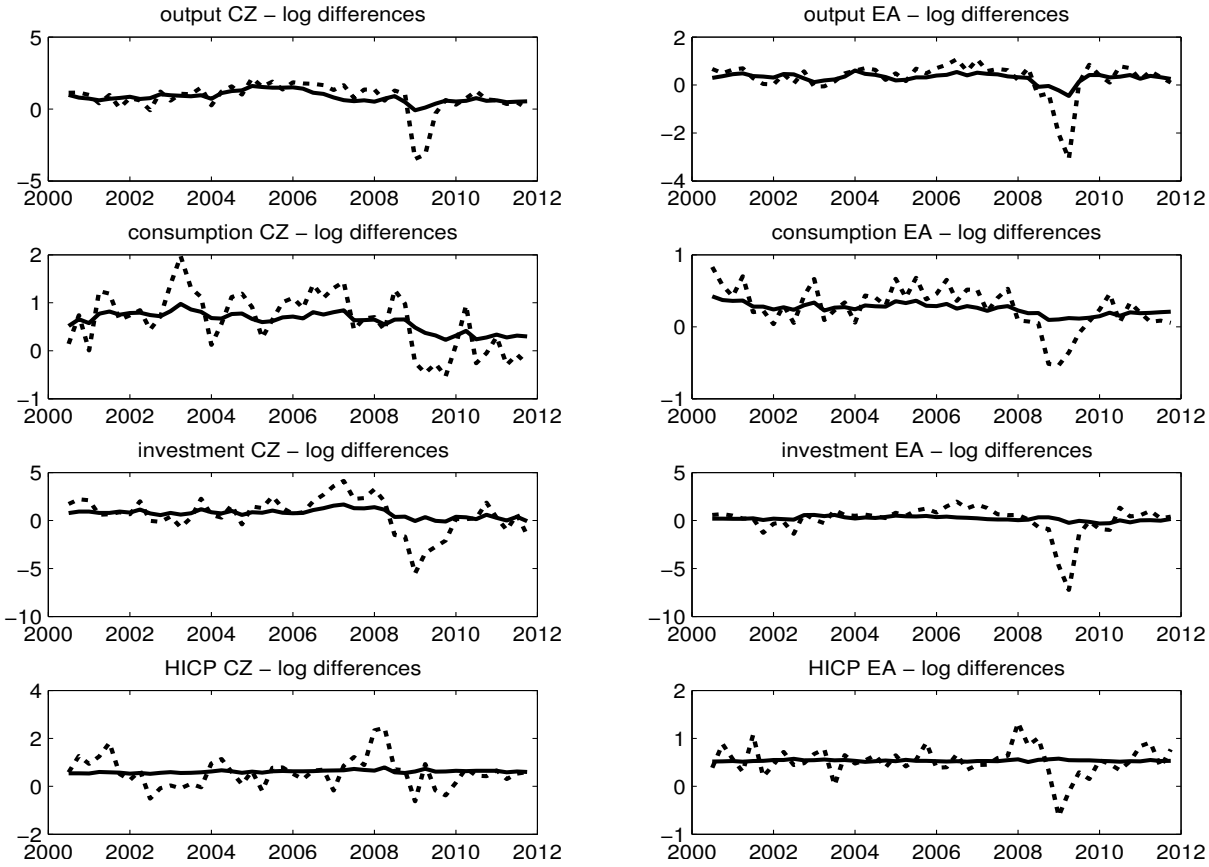


Figure 1 Original Data - dotted line, Filtered Trend - solid line

[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.

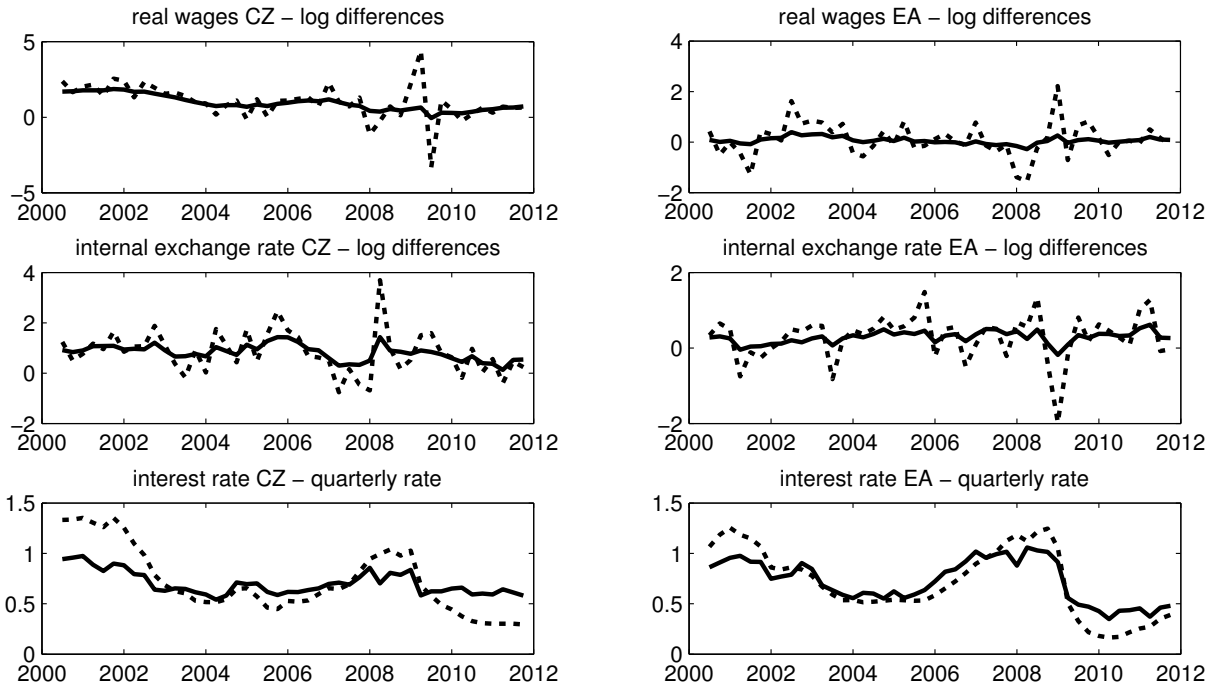


Figure 2 Original Data - dotted line, Filtered Trend - solid line

It is worth mentioning that according to the Bayes factor (BF=186.54) the proposed specification decisively outperforms the original specification of the model estimated on the demeaned log differences with the same number of shocks. Figures 1 and 2 display original data and the filtered trends. Results suggest that this method is able to find very reasonable trend in the data. We can see that while for some variables, e.g. investment and prices, the filtered trend is very similar to the mean of the observable, for some variables, especially for the interest rates, the filtered trend is more fluctuating. However, this result is based on the employed calibration and can be easily modified by the different calibration of the standard deviations of trend shocks. For the sake of comparability I decided to calibrate all trend shocks equally to one third of the standard deviation of the observable, however, it is possible to calibrate the trend shocks differently for each variable. This feature of the proposed trend specification can be interpreted such that this specification allows researcher to decide how much portion of the data he/she wants to explain by the trend shocks and consequently the model itself decides about the distribution of these shocks in the data.

Figure 3 HP output gap - dotted line, model output gap - solid line

As regards the economic implications of the model, we can compare the output gap implied by the model with the output gap obtained by HP filter, see Figure 3. We can see that the model output gap and the output gap obtained by HP filter are pretty similar, however, there are some remarkable differences. The model output gap implies much milder recessions in 2002 - 2005 and 2009-2011 and much stronger boom in 2005 - 2008 than the output gap obtained by HP filter. Unlike the model output gap, output gap obtained by HP filter also implies that the recessionary gap of EA 12 in the current crises is closed, which I regard as a very odd result. This can be a consequence of the well known end-point bias of the HP filter. In my view the model output gap provides more plausible description of the business cycle position of both economies in current crisis.

6 Conclusion

In this paper I propose flexible trend specification for estimating DSGE models on the log differences. I demonstrate this flexible trend specification on a New Keynesian DSGE model of two economies, presented in Kolasa [11]. The advantage of the proposed trend specification is that the trend component and the cyclical component are modeled jointly in one model and the model itself decides which part of the data belongs to the trend component and which part belongs to the cyclical component. The proposed trend specification is flexible in the sense that smoothness of the trend can be easily modified by different calibration of the trend parameters. Results suggest that this method is able to find very reasonable trend in the data. Moreover, according to the Bayes factor the proposed specification decisively outperforms the original specification of the model estimated on the demeaned log differences with the same number of shocks.

Acknowledgements

This work is supported by the Research Centre for the Competitiveness of the Czech Economy. I also thank Tomáš Motl for his useful comments.

References

[1] Adolfson, M., Laseen, S., Linde, J., Villani, M.: Bayesian Estimation of an Open Economy DSGE Model with Incomplete Pass-Through, *Journal of International Economics* **72(2)** (1973), 481–511.

[2] Aguiar, M., Gopinath, G.: Emerging Market Business Cycles: The Cycle Is the Trend, *Journal of Political Economy* **115(1)** (2007), 69–102.

[3] Andrle, M.: The Role of Trends and Detrending in DSGE Models. MPRA Paper No. 13289, (2007).

[4] Brůha, J.: An Empirical Small Labor Market Model for the Czech Economy, *Czech Journal of Economics and Finance* **61(5)** (2011), 434–449.

[5] Burstein, A. T., Neves, J. C., Rebelo, S.: Distribution Costs and Real Exchange Rate Dynamics during Exchange-rate-based Stabilizations, *Journal of Monetary Economics* **50(6)** (2003), 1189–1214.

[6] Calvo, G.: Staggered Prices in a Utility Maximizing Framework. *Journal of Monetary Economics*, **12(3)** (1983), 383-398.

[7] Canova, F.: Detrending and Business Cycle Facts, *Journal of Monetary Economics* **41(3)** (1998), 475–512.

[8] Canova, F.: Bridging Cyclical DSGE Models and the Raw Data. Unpublished manuscript, (2009). Available at http://www.crei.cat/people/canova/

[9] Canova, F., Ferroni, F.: Multiple Filtering Devices for the Estimation of Cyclical DSGE Models, *Quantitative Economics* **2(1)** (2011), 73–98.

[10] Cogley, T.: Estimating and Testing Rational Expectations Models when the Trend Specification Is Uncertain, *Journal of Economic Dynamics and Control* **25(10)** (2001), 1485–1525.

[11] 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.

[12] Koop, G.:*Bayesian Econometrics*. John Wiley & Sons Ltd, Chichester, 2003.

[13] McGrattan, E. R.: Measurement with Minimal Theory, *Quarterly Review* **33(1)** (2010), 2–13.

[14] Smets, F., Wouters, R.: An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area, *Journal of the European Economic Association* **1(5)** (2003), 1123–1175.

[15] Whelan, K.: New Evidence on Balanced Growth, Stochastic Trends, and Economic Fluctuations. MPRA Paper No. 5910, (2006).

Classification of the electronic retail core banking market consumers

Ivan Soukal[1], Martina Hedvicakova[2]

Abstract. This paper is focused on the retail core banking services market. In the Czech Republic the consumers can use web accessible free of charge comparison tool for current account offer called the retail bank charges calculator. This system's database holds day-to-day usage patterns of every consumer using this calculator. The aim is to classify the Czech e-banking clients that used the calculator during the year 2011. As a data preparations there were performed verification-validation phase, transformation and dimension reduction. The two-step cluster analysis was performed on 11 255 records classified by 19 variables regarding the type of the service, moth usage frequency or average amount of money, communication channel used to establish or order the service and the criterion if the service falls within or outside the single provider. Analysis indentified as an optimal number of clusters 3. There were identified mainstream client, active client and the client with the mixed communication channel preference.

Keywords: consumer, cluster, retail core banking services, calculator.

JEL Classification: C38, D12
AMS Classification: 91C20

1 Introduction

It is not a simple task to get the market overview in the retail core banking services (thereinafter abbreviated as RCBS) offer. The opportunity costs are high and so the Retail bank charges calculator system (thereinafter abbreviated as Calculator) was created as a web accessible free of charge comparison tool for current account offer. The database of the Calculator now holds the tariff data of 13 banks (more that 98% of the RCBS market in the Czech Republic) and their current accounts. Consumer can use the Calculator to gain individualized market price overview. The collateral benefit to information asymmetry reduction on the RCBS is the dataset of consumer usage patterns.

This dataset will be used to classify the consumers however the users of the Calculator cannot be considered as the adequate sample from the Czech Republic RCBS consumer population. Still the data are the description of the major group of consumers (consumer s with activated electronic banking) and so it can be used for marketing or research purposes. The goal of the paper is to identify these usage patterns regarding the automated teller machine (thereinafter abbreviated as ATM), direct payments, standing orders, encashment, turnover or account balance.

2 Methodology

2.1 The goal and previous problems

The aim of the analysis is to classify the electronic retail core banking services consumers in the Czech Republic into the basic groups by the usage patterns. This paper continues and complements the effort from. [6] There the k-means method was applied to classify the respondent answers gathered during the pilot run of the Calculator and so we demonstrated one of the possible utilization of the Calculator output. The number of clusters determination was based on G5 criterion computation. [4] One of the problems, that occurred, was the asymmetry between the number of transactions and concentration of low frequency of transactions and low turnover consumers. This asymmetry resulted in the computation of one main, or there can be said "mainstream", cluster that contained almost 70 % of the analyzed members. The rest consisted of several smaller, less significant clusters mainly containing consumers showing extremely high numbers of transactions and high turnover on the current account. Part of the problem was wide interval of allowed values, respectively usage frequencies and inclusion

[1] University of Hradec Králové, Faculty of Informatics and Management, Rokitanského 62, Hradec Králové 500 03, Czech Republic, ivan.soukal@uhk.cz.
[2] University of Hradec Králové, Faculty of Informatics and Management, Rokitanského 62, Hradec Králové 500 03, Czech Republic, martina.hedvicakova@uhk.cz.

of consumers that do not have electronic banking activated. However those facts did not caused the large mainstream cluster issue.

2.2 Data source

Data gathering process is conducted by the Calculator's respondent database that can be exported to .csv file. In this analysis the respondent for the year 2011 were analyzed. The Calculator's GUI for consumers can be found at bankovnipoplatky.com server. The consumer enters his or hers usage of the RCBS into the online form and then the individual market price overview is computed and displayed. Therefore it is necessary to presume that the findings of our analysis are limited to consumers with access to the Internet and possessing at least basic level of ICT literacy. Furthermore only the consumers with PC access to account were analyzed (regular reports of [2] show that the share of these accounts is rising and we presume that now it is more than 65 %). Also there has to mentioned that the form for the consumers does not verify the uniqueness of the respondents and so we cannot avoid repeated access of the same respondent, however our number of respondents is high enough to set this potential problem aside. For more information about the Calculator, please see [6] or [7].When the Calculator's form is filled and the price commutation starts, the inputted figures are saved. From the marketing research point of view there are gathered data:

- Multivariate – there has been monitored 53 variable concerning RCBS usage, 2 system variables for respondent identification and 45 variables containing the calculated costs for each of monitored RCBS product,
- Primary – data were gathered directly from the client,
- Subjective – data came from respondent himself, respectively it is his or hers subjective seem.

It is clear that our sample cannot represent the entire RCBS consumer population in the Czech Republic however the Calculator's database holds unique data set that is practically impossible to obtain by usual market survey methods due to the high costs. Moreover there has to be taken into the consideration the motivation, to be more specific the value added. The market overview is highly personalized and so it is in the best interest of the consumer to fill in the form as accurate figures and answers as possible, otherwise the Calculator's output will not help at all.

2.3 Data preparation

Data gathered using the online forms contain a number of faults. Even though the form was improved since the pilot run, yet it fails in data validation e.g. data include obvious errors, blank values or wrong data type. The data preparation was necessary to:

1. decide what variables are suitable for the analysis – due to wide range of the monitored services there has to be determined what to include into the analysis. Considering rarely used services it can be assumed that such services characterize the consumer only superficially and the use of such variables in the cluster analysis could result in the creation of artificial clusters, where the cause criterion would be those services. For each service, respectively variable, the frequency table and mean are computed and evaluated.
2. remove the records with invalid values – number of records in the database contained blank or invalid values. At first there were marked for removal the records containing text values in numeric variables. The correction in some cases is possible. Nevertheless this loss in the file does not mean a significant change in the structure of the clients. Then the interval of illegal values is set because the analysis in [6] confirmed the hypothesis that RCBS products can be also used for business purposes. This is not only in violation of the focus of current research, but mainly in violation of retail bank–client contract. With this in mind frequency tables were again checked and individually assessed for the cut-off frequency.
3. decide the missing values treatment – the literature is usually recommending:

a) replacement by the mean value of the variable,
b) generating a random number from the distribution of the variable,
c) using predicted values from regression model based on other variables,
d) delete the whole record. [3]

In our specific case the variables should be assessed individually with regard to their role and significance in the model. It has to be taken into account the fact that missing values in the questionnaire might have different reasons. Therefore it is necessary to distinguish whether the value is blank because the client does not use the service, or whether the value is blank due to omission or otherwise. That is why it was chosen to treat missing values by:

a) the replacement value of zero – although this operation is not a standard one, in this case it is tenable. It is expected that if the client uses for direct payments telephone and electronic banking, then for the same service the "at the desk" channel is not demanded. So we presume for certain variables that blank values mean: "I do not use.". Therefore the missing values should be replaced by zero. Some services in the cal-

culator, [6] or [7], are obviously complementary services, where the task of filling in all the fields, would only hinder the consumer.

b) the regression model – not all variables can be substituted by zero. There it is used the regression model where independent variable will be determined from the correlation matrix and predicted value is saved. Using multivariate regression in our case is inappropriate.

c) exclusion of the record – in case the number filled in values is insufficient, the record is removed from the file.

2.4 Data transformation

As mentioned earlier, the analysis published in [6] showed problem of clusters characterized by extreme values and very large mainstream cluster. Use of appropriate transformation will help the symmetrization of the distribution and so to suppress importance of high values of individual variables and to be able of more detailed assessment of the major cluster. Certain number of the variables showed single peak distribution skewed to left. For these variables a logarithmic transformation showed to suitable:

$$X^* = \ln(x+1) \tag{1}$$

2.5 Multicollinearity elimination

Due to the nature and number of analyzed variables can be expected interdependence of the variables. Multicollinearity could significantly affect the output of clustering. To eliminate multicollinearity there was used the principal component analysis method. This method computes a new set of linearly uncorrelated variables from the set of possibly correlated, while reducing the dimension of vector space. The analysis is in the IBM 18PASW (formerly SPSS) included the factor analysis.

There cannot be presumed that all variables have the same weight because the variability differs greatly, there was chosen the extraction based on correlation matrix. Using the correlation matrix extraction usually makes harder to interpret the component, however in this case the interpretation is not needed at all. When determining the optimal number of new variables (there was used non-rotated solution), there can be followed several rules and recommendations [4], [5] such as:

- the value of eigenvalues should be 1.0 and higher however the interval of eigenvalues (0,7; 1) can be consider if revision needed,
- described cumulative variability should be at least 70 % of the original,
- the number of components should be significantly lower than the number of original variables.

Considering our data collection gathered using the questionnaire survey, there can be assumed higher number of components, or a lower level of cumulative variability described. Principal component analysis also solves the problem of scale. In our case there is used its transformation function where e.g. the encashment has naturally lower frequency (mainly considering SIPO payment the bank considers as one payment with no regard to how many orders are made under one SIPO number) than direct payment usage and extremely lower considering significantly different scale of month turnover and average balance in €.

2.6 Cluster analysis

The aim of this paper is to determine the major consumer groups by their usage patterns – what and how often they demand. Given this goal there has been chosen the method of cluster analysis. Compared to the [6] where the k-means algorithm was used, here we used two-step cluster analysis that allows using log-likelihood or euclidean distance measurement (K-means is Euclidean based only).

The pre-cluster step uses a sequential clustering approach. It scans the data records one by one and decides if the current record should be merged with the previously formed clusters or starts a new cluster based on the distance criterion. The procedure is implemented by constructing a modified cluster feature (thereinafter abbreviated as CF) tree, for CF tree and BIRCH algorithm details see [8]. This method allows working with large datasets because using CF characteristics that represent separated data groups this method abstract from single records (cases) inside the group. This reduces computation time and resources.

In two-step cluster analysis we decided to use log-likelihood distance measurement, based on the decrease in the likelihood function when merging clusters. This method was introduced in [1]. The algorithm can estimate the optimal number of clusters by use one of BIC (Bayessian's Information Criterion) or AIC (Akaike's Information Criterion). The BIC characteristic was used. The BIC for each number of clusters within a specified range is calculated and used to find the initial estimate for the number of clusters. Then the initial estimate is

refined by finding the largest relative increase in distance between the two closest clusters in each hierarchical clustering stage. The BIC is described e.g. in [4]. Noise reduction was not used because of the risk of possible suppression, although small but valid cluster of clients preferring a e-banking however using the desk services too.

3 Analysis

3.1 Raw data treatment

There were declared types and scale of the variables and so the interval of valid values. Then there was analyzed which variables will be taken into consideration for the upcoming analysis (descriptives and frequency of missing values). All qualitative variables such as card type, homebank, the type of statement etc. were excluded. From qualitative variables there were excluded communication channels of telebanking and collection box because of very low usage and the same case was cash-back service, deposit and withdrawal of high amount of cash, ATM usage abroad. Total of 21 variables was chosen for further analysis, for the list of the variables please see tab. 1 at the end analysis chapter.

The choice of variables was followed by the recode of missing values but in chosen variables only, please see the methodology. Regarding the specific methods of recoding the missing values by zero, there has to be mentioned that in previous research [6] there was identified that most respondents are interested in only less than half the questions, meaning the empty field can be without a doubt considered as unused service. For the variable of average turnover and minimal balance no recode was performed. The variable "ATM withdrawal amount, other bank" there was used the regression of function with independent variable "ATM withdrawal amount, other bank" (the rest of the variable showed only low correlation from -0,18 to 0,23 value). At the end of this phase there was conducted the search for the records service relation errors e.g. when no card selected, then no ATM withdrawal can be made. After this phase there was left 11 255 records that were taken into the next analysis by listwise missing (missing value is blank value or used defined illegal value) value treatment. It lowered the initial number of records from 17 402 to 11 255. It seems as a high number of lost records however there has to be taken into consideration very specific form of data acquisition via the Internet.

3.2 Transformation and dimension reduction

After the verification/validation phase the 21 selected variables was transformed using the formula 1 and saved as a new ones. Due to high dimensionality and the presumption of the cluster analysis of uncorrelated variables there has to be treated the possible interdependence. In the correlation matrix there was found that there is correlation between some variables, to be more specific Direct payments to other bank at desk – Encashment to other bank at desk, Incoming payment from other bank – Incoming payment from own bank. Those correlation coefficients were around 0,45 value. That is why we decided to use principal component analysis.

On the scree plot chart the 4th and 6th components were obvious elbows but were followed by components with eigenvalue higher than one and moreover the total variance explained was lower than 70 %. The last component with eigenvalue higher than one is 8th one and the total variance explained in non-rotated solution is almost 70 and it can be considered almost as an "elbow" point by sight. Still there has to mentioned that both sources [4], [5] that scree plot interpretation is very individual. New variables were saved and were used for the last phase of the analysis.

3.3 Cluster analysis

Due to high number of records only nonhierarchical or modified hierarchical methods were taken into consideration. We decided to use two-step cluster analysis, please see methodology. 8 components were used as continuous variables for clustering using log-likelihood metric and BIC optimal cluster criterion. The CF tree was extended to 5 levels because initial 3 level settings showed unsatisfactory results.

There was computed optimal number of clusters 3. For each cluster there was computed the centroid in initial scale for analyzed variables. There was computed also 4 cluster analysis however the next cluster did not show enough of interpretation value. For 3 cluster computation output see the table below, all frequencies of usage are per mensum, own/other bank reflects that payments can be made between the accounts provided by the same bank or by two independent providers. Amount of money are converted by exchange rate EUR/CZK = 24,29 where EUR is the base currency.

Variable/cluster	**1**	**2**	**3**
Relative cluster size in %	8,0	34,7	57,3

Minimum account turnover in €	709	838	801
Average account balance in €	1 108	1 005	1 098
Domestic ATM withdrawal, own bank	3,34	3,13	2,90
Domestic ATM withdrawal, other bank	0,67	1,85	0
Domestic ATM withdrawal amount, other bank	31,09	94,25	0
Incoming payment from other bank	2,39	2,67	2,37
Incoming payment from own bank	1,26	1,19	1,00
Direct payments to own bank at desk	0,77	0	0
Direct payments to own bank Internet	1,96	2,24	1,98
Direct payments to other bank at desk	0,96	0	0
Direct payments to other bank Internet	2,79	4,54	3,96
Standing orders to own bank at desk	0,62	0	0
Standing orders to own bank Internet	0,97	1,15	1,05
Standing orders to other bank at desk	1,13	0	0
Standing orders to other bank Internet	1,33	2,72	2,53
Encashment to own bank at desk	0,37	0	0
Encashment to own bank Internet	0,40	0,44	0,41
Encashment to other bank at desk	0,56	0	0
Encashment to other bank Internet	0,55	1,02	0,98
Cash deposit at desk	0,69	0,44	0,29
Cash withdrawal at desk	0,54	0,29	0,09

Table 1 average RCBS usage of indentified clusters

In the table there are three usage patterns that can be identified as:

1. mixed preference consumer – even this consumer has active electronic banking, he or she prefers for mostly standing orders and about one third of direct payments at the desk service. Due to higher price of these services it is unusual however the share is low and so generally we can presume that most of consumers in analyzed population prefer strongly e-banking. Considering the overall activity of 4 ATM withdrawals, more than 3 incoming payments, more than 6 direct payments and 6 regular payments it is slightly above average consumer mainly in ATM usage confirming certain preference of cash (and cash services).
2. active e-banking preferring consumer – the third of the analyzed sample is more active then the major one but shares almost total preference of electronic banking. Overall activity varies from 40 % above average to 10 %, 5 ATM withdrawals without regards to the ATM owner, almost 4 incoming payments, almost 7 direct payments and more than 5 regular payments.
3. average consumer – almost total preference of electronic banking and compared to the rest also of ATM of own provider where withdrawals from the other provider ATM is more charged. Average activity is then 3 ATM withdrawals slightly more than 3 incoming payments, 6 direct payments and 5 regular payments.

The analysis also showed that turnover of more active consumer is not significantly higher meaning without noticeable interpretation value. If e.g. the ANOVA would be performed, than due to high number of records statistical significance would be likely found still interpretation value is almost none. This is surprising and it deserves further analysis because it is expectable dependence we are about to analyze in future.

4 Conclusion

The analysis proved that there can be identified specific usage patterns in the dataset of the respondent answers in the Calculator. There were identified 3 clusters in the sample of the electronic banking activated consumers with Internet access to account and basic ICT literacy: mixed preference consumer, active consumer and the average consumer. The difference between the clusters is not considering just the frequency of usage but also the range of used services. The mixed preference consumer uses the branch services however the electronic banking is activated. The average client differs from the active one not just by the frequency but also by the preference of ATMs of own provider. These usage patterns can be used in the further research not just to describe the demand side of the market but also to be the base of the price register. The consumer would just identify the closest cluster. For each cluster the bank would have to state the price without the product tying but fixed one.

Acknowledgements

This paper is written in the frame of specific research project *"Vyspělé technologie pro podporu efektivního podnikového řízení a bankovnictví"*, translated as *"Advanced technologies for effective business management support and banking"*, project order 2108 on the University of Hradec Králové Faculty of Informatics and management, funded by Czech Republic Ministry of Education, Youth and Sport.

References

[1] Chiu, T. D.: A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment. In *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining*. San Francisco, 2001.

[2] Czech national bank. *Financial market supervision reports*. [online]. 2010, [cit. 2012-04-01]. From Financial market supervision reports - Czech National Bank: http://www.cnb.cz/miranda2/export/sites/www.cnb.cz/cs/dohled_financni_trh/souhrnne_informace_fin_trhy/zpravy_o_vykonu_dohledu/download/dnft_2010_en.pdf.

[3] Hebák, P., Hustopecký, J., Jarošová, E., Pecáková, I.: *Vícerozměrné statistické metody. vol. 1.* Informatorium, Prague, 2004.

[4] Hebák, P., Hustopecký, J., et Pecáková, I. [eds.]: *Vícerozměrné statistické metody. vol. 3.* Informatorium, Prague, 2005.

[5] Meloun, M., Militký, J., Hill, M.: *Počítačová analýza vícerozměrných dat v příkladech.* Academia, Prague, 2005.

[6] Soukal, I., Hedvičáková, M.: Retail core banking services e-banking client cluster identification. *Procedia Computer Science Journal* **3** (2010), 1205–1210.

[7] Soukal, I., Hedvičáková, M.: Retail core banking services costs optimization. *Procedia Computer Science Journal* In print

[8] Zhang, T., Ramakrishan, R., Livny, M. Birch, R.: An Efficient Data Clustering Method for Very Large Databases, In *Proceedings of the ACM SIGMOD Conference on Management of Data*. Montreal, 1996.

Mathematical model of economics of municipal waste management

Jana Soukopová[1], Jiří Kalina[2]

Abstract. The paper discusses the mathematical and economical model of municipal integrated waste management system (IWMS) that has been developed for the real needs of the decision support of the Ministry of Environment of the Czech Republic and regional governments. The model of IWMS is designed as universal and is implemented as a combination of three models including environmental and economic point of view. The model allows evaluation of the economic and productive efficiency of the system due to own setting of input values and can be used for waste management planning as a decision support tool. Model involves composting, energy and material recovery of waste and landfilling. Its size (number of sources and facilities) depends only upon available data.
The model was developed based on given input macroeconomic variables and it enables to inclusion or exclusion of certain equipment of waste management and the capacity of the equipment. It uses data from annual reports of municipalities (if available) on the production of municipal solid waste and estimates its quantity (if unavailable) by using a sophisticated model, including demographic and socio-economic impacts. The important component of the paper is economic model of the IWMS functions.
Keywords: waste management, model, system.

JEL Classification: C690
AMS Classification: 91B99

1 Integrated waste management system in the Czech Republic

In order to define the term waste management, we should first define what waste is. In Chapter I, Article 3 of Waste Framework Directive – Directive 2008/98/EC [17] (from now just Directive), **waste** is defined as *"any substance or object which the holder discards or intends or is required to discard"*. Czech Act no. 185/2001 Coll. (Waste Act) [16], Part 1, Section 3 adds one more condition in order for this substance or object to become waste – it has to be *"specified in some of the waste categories stipulated in Annex 1 to Waste Act"*. In the Directive we can find also definition of waste management which is *"the collection, transport, recovery and disposal of waste, including the supervision of such operations and the after-care of disposal sites, and including actions taken as a dealer or broker"*. Other terms that will be used throughout this thesis are waste collection and waste treatment. The Directive defines these terms as *"the gathering of waste, including the preliminary sorting and preliminary storage of waste for the purposes of transport to a waste treatment facility"*, and *"recovery or disposal operations, including preparation prior to recovery or disposal"*. The last term to be defined here is municipal waste (MW) and municipal solid waste (MSW). The Directive does not specify what the municipal waste is (however, the Directive operates with this term), therefore we need to look for the definition in the Czech Waste Act, Section 4, where the municipal waste is defined as *"all waste generated in the territory of a municipality in connection with activities of legal entities or natural persons and which is stipulated as municipal waste in the statutory instrument, with except of waste produced by legal entities or natural persons authorized"*. As Hřebíček et al. [3] notes, definitions of these terms are in the Directive and Czech Waste Act slightly different in details but for the purpose of this thesis it irrelevant.

The decisions in the area of MW management are not only capital intensive, but also difficult from environmental and social points of view. There is the need to develop, master and implement simple but reliable mathematical-economic model that will help to the decision in the analysis of waste management processes. The first works on models of waste management costs and prices and their optimization (in terms of efficiency) go back to the 60s and 70s. In that time the model of waste management was seen only through the terms waste collection and disposal (on landfills).

[1] Mgr. Ing. Jana Soukopová, Ph.D., Masaryk University, Faculty of Economics and Administration, Department of Public Economy, Lipová 41a, 602 00 Brno, soukopova@econ.muni.cz.
[2] Mgr. Jiří Kalina, Masaryk University, Faculty of Science, Research Centre for Toxic Compounds in the Environment, Kamenice 126/3, 625 00 Brno, kalina@recetox.muni.cz.

This paper discuses a new model of integrated waste management system (IWMS) to assist in identifying alternative solid waste management strategies and plans that meet cost, material, energy, and environmental emissions objectives.

The MSW is all waste generated within the community (cities and villages) by the activities of its inhabitants (households) and businesses (e.g. trade waste), which is separated into its components and transported to waste treatment facilities where is recovered or disposed. The MSW normally contains the remains of food and vegetables, paper, plastic, glass and metal containers, printed matter (newspapers, magazines, and books), destroyed products, ashes and rubbish, used or unwanted consumer goods, including shoes and clothing. The MSW (or its separated components) can be composted, used as raw material (paper, plastic, glass, and metals), used in biogas, energy recovery (incineration) plants or land-filled. The separation of its components may take place at the source (separate collection in the municipalities) or in the facilities. We analyses the post-consumption stages of the waste life cycle, namely collection, sorting, treatment and final disposal. The IWMS is illustrated by Fig. 1 of Shmelev and Powell [11] which shows the main material flows within the system, emissions, etc.

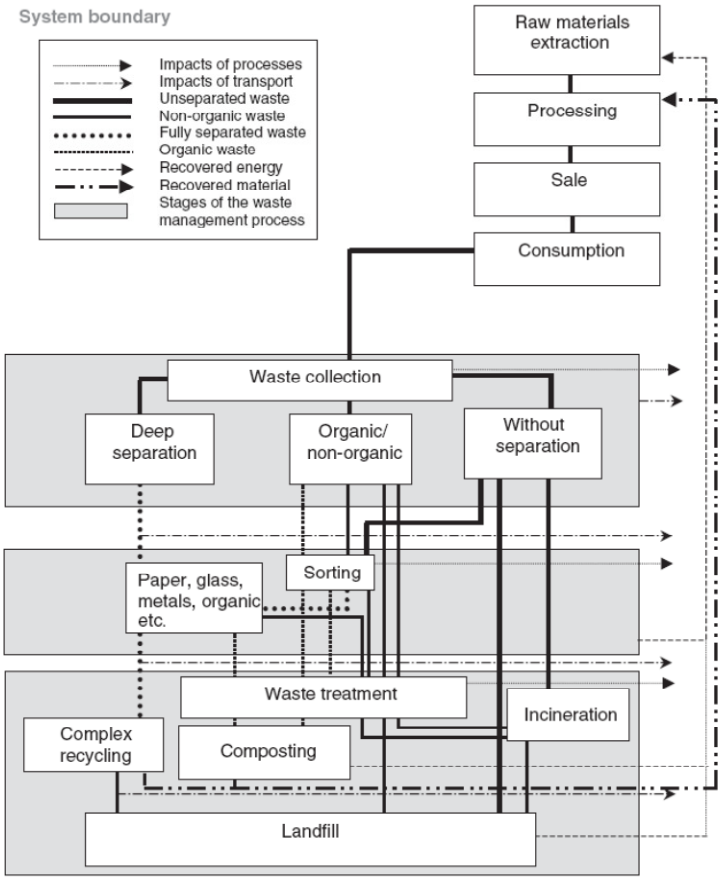


Figure 1 The IWMS: material flows [11]

The Fig. 1 reveals that the whole life cycle of materials entering and leaving the waste management system consists of several stages (raw materials extraction, processing, sale, consumption, finally becoming waste when they are discarded by consumers. These materials in the waste stream then undergo collection, sorting (removal of recyclable materials) and treatment (which can be thermal or biological) with the final stage being disposal in the landfill. We can define the individual waste streams which are mass balancing. The shaded areas in the Fig. 1 are the stages of the life cycle of MSW taken into account in this paper and we simplified these to waste streams between producers and treatment facilities including transport.

2 Mathematical model of economics of IWMS

Earlier this decade, the development of mathematical-economic models of WM has moved towards the *integrated model of waste management* (IMWM), which is designed to minimize the economic costs and / or environmental impacts, see Berger et al. [1], Wang [14], Yeomans [15]. It already requires the use of optimization

procedures for finding the minimum of appropriately defined objective function (total cost, emissions, etc.), Haigh [2].

Consider the IMWM discussed by Hřebíček et al. [3] and Hřebíček and Soukopová [5,6] which consists of the set of MSW sources (municipalities) of the Czech Republic connected by the road network with the set of waste treatment facilities (composting, biogas, mechanical-biological treatment and pre-treatment of recyclable waste plants, incinerating plants with energy recovery and landfills) where MSW (or its components separated at source) is transported to chosen facilities for recovery or final disposal. The material balance is examined in terms of material flows between MSW sources and waste treatment facilities. The waste treatment facility technologies depend on both the operators (voluntary cooperation, market) and the regulator (government). The regulator is able to use the operating permits or economic instruments (charges) so that the waste management will show a minimal impact on the environment in socially viable cost for the most of communities.

In developing IMWM of the Czech Republic, we came out of the mathematical-economic models available in literature. Since the early 90 years, a number of IMWM has been developed which were based on life cycle analysis (LCA), i.e. materials and energy balances, see McDougall et al. [9, 10] and Solano et al. [12]. Most available models are static, respectively deterministic and quantify the uncertainty of estimates due to random nature of input values. Another disadvantage of models based only on the LCA is that they do not allow optimizing the allocation of waste treatment facilities from sources and / or quantifying the transport emissions. We tried to reduce the greatest uncertainty of our model by the estimation of the composition of municipal waste, waste separation, varying the proportion of resources, varying quantities of trade waste and the like.

The developed IMWM of the Czech Republic consists of the combination of three sub-models where we used following tools [5, 6]:

1. The geographic information system (GIS) *ArcMap,* which computed a transport matrix linking the sources MSW and waste treatment facilities and the simple model which generated emissions from the transport of MSW and enable to find the closest facility.
2. The cost economic sub-models of every type of waste treatment.
3. The regression analysis model for the determination of the quantity and composition of MSW. Model based on an older version [4, 5] estimates the waste production at every source (municipality) using data from a reference sample of municipalities with known level of production or based on annual municipal waste reports of the quantity of individual MSW components.

The above IMWM requires criteria (prioritization) from decision makers (regulators), which may involve an acceptable level of pollutant emissions and costs, as well as a reduction of landscape and biodiversity or prevent a pollution of groundwater and surface water. Practically, such optimization comes into the consideration of regulators (government) when deciding on localisation of new facility (technology and capacity) and / or closure of existing facilities, the regulation of their capacities and so on. A chosen feasible minimum is usually acceptable for regulator without optimization. It should be used only for the Environmental Impact Assessment (EIA) of a new facility (assessing alternatives) but also in the Strategic Impact Assessment (SEA) of strategic documents such as plans for regional development or waste management plans at the county level.

2.1 Mathematical model of quantity and composition of MSW

Predictive model of quantity and composition of municipal waste is based on an earlier published model [4, 5]. In the first phase of modelling, the general form of the function f_p describing dependence of generated municipal waste from municipalities on selected parameters is stated. In the second phase up to 8 different coefficients of this function are calculated and in the third phase the model allows to estimate the quantities of individual waste types produced by municipal waste from the knowledge about the parameters of each municipality.

The model works by regression analysis method and allows to continuously improving the accuracy of estimates of production by setting of internal calibration constants and by choosing of different products of the input parameters. That is, why the model is in constant evolution. Detailed description of all functions of the model is not possible given the extent of the paper, so the following text will focus on description of a default method, which is available in the model and was used for determination of the production in the municipalities.

For the estimate 14 input parameters for each municipality are used. By using automated wrapper the model retrieves values of the parameters from the interface of web based information system of Regional Information Services. The parameters for which is expected possible binding to waste production were chosen: population *inh*, number of retired citizens *pens*, unemployment *unem*, gasification *gas*, cadastral acreage *cad*, acreage of public lawns *gras*, acreage of public and private gardens *gard*, recalculated number of schools *scho*, number of hospital facilities *hosp*, number of businesses companies *bus*, longitude *long*, latitude *lat*, altitude *alt*, and status of the municipality *stat*.

After assigning numerical values to parameters *gas*, *scho* and *stat*, which are not originally in numeric format, and addition of constant values to the parameters *unem*, *pens*, *gas*, *long*, *lat*, *alt*, which are not assumed to have a direct correlation with appropriate production of waste (i. e. production is not zero when these parameters decrease to zero) the production function f_p can be defined as follows:

$$f_p=(c_1+inh\cdot pens\cdot unem\cdot gas+c_2\cdot cad+c_3\cdot gras+c_4\cdot gard+c_5\cdot scho+c_6\cdot hosp+c_7\cdot bus)\cdot(c_8+long\cdot lat\cdot alt\cdot stat), \quad (1)$$

where c_i, $i=1,..,8$ are searched constants of the model.

After multiplying, the expression contains 16 summands consisting of multiples of the original parameters (also called properties) of the formula (1) that are entering the regression analysis of the reference sample of municipalities, whose production is known:

1. The known production is expressed as a sum of the f_p function value and estimation errors e_j for each municipality.
2. Total relative[3] square[4] error of estimate can be expressed as $E_p = \sum_{j=1}^{n} \left(\frac{e_j}{inh}\right)^2$, where j goes over the all municipalities in the reference sample.
3. The expression for E_p is partially derived stepwise with respect all properties and the corresponding values are transposed into a $m \times m$ matrix.
4. Searching for the minimum total square error corresponds to the situation, when the first derivatives of the error by all the properties are equal to zero. The optimum values of the searched constants can thus be found as a solution of the matrix-expressed set of equations, with right hand side represented by column vector, where each number is a sum of products of the production of individual municipalities and properties appropriate to the matrix line, divided by the number of inhabitants[3].

The calculated coefficients are saved in the profile with the specified name for later use and can be used for the estimate of the production of various waste types for all municipalities for which are available data from the Regional Information Services.

To evaluate the accuracy of the selected profile, the model provides the information about deviations of the result estimate made by regression analysis compared to the known production of the sample of municipalities. If parameters are properly selected, can be achieved accuracy of 10%.

2.2 Transport cost network model

Consider the MSW flows at the Czech Republic among all sources (municipalities) S_i, ($i = 1\ldots n$), $n = 6245$ and all waste treatment facilities F_j, ($j = 1\ldots m$), $m = 471$, where:

$$m = MC + MB + MT + MI + ML,$$

where

MC means number of composting facilities (236 in 2012 year);
MB means number of biogas plants (9 in 2012 year);
MT means number of mechanical-biological treatment and pre-treatment of recyclable waste plants;
MI means number of incinerators (with energy recovery – 3 in 2012 year);
ML means number of landfills (223 in 2012 year).

Consider these MSW flows in a continuous manner and mass balance between sources and facilities carry out over a longer period of time (annual reporting). If we modelled the allocation of existing n sources S_i and m facilities F_j in the Czech Republic, then we built the transport matrix $\boldsymbol{D} = \{d_{ij}\}$, ($nxm$), of real transport distances d_{ij} (e.g. road maps) among all sources S_i and all facilities F_j and the vector of the distance $\boldsymbol{dc} = \{dc_i\}$ (nx*1*) of the source S_i from its closest landfill F_c, $c \in \{1,\ldots,m\}$ [5], [6].

2.3 Mathematical model of economics of facilities

We developed mathematical model of economics (costs) for all types of facilities F_j (j=1…m), i.e. composting, biogas plants, mechanical biological treatment (MBT) plants, incineration plants with energy recovery (ERP) and landfills. These models are similar and we introduced this economic model for a generic facility *F*.

The price function is the function of these variables:

[3] The size of the production determination errors is related to the total number of inhabitants, because the estimate error in the order of tens of tons is negligible in the case of a large city, but fatal in case of a small village.
[4] Minimisation of the square error is the same as minimisation of the absolute value of the error.

$$p = f(B, C, I, K, T, u, l, r, N), \tag{2}$$

where

B is the total revenue generated from the facility;
C is the total operating costs arising from the facility;
I is the investment expenditures in the facility;
K means the capacity of the facility;
T means a tax on income;
u means the interest due on loans;
j means repayment of principal on loans;
r means the discount rate;
N means the lifetime of facility.

Calculate the price p of one ton of the waste treatment at a new composting, biogas, MBT and ERP plant F. This calculation is based on the financial and economic analysis and financing methods for the measuring the efficiency of investment, see Levy and Sarnat [8], Valach [13], etc.

We used the *Net Present Value* (*NPV)* as the basic calculation method for the price p.

$$NPV = -I + \sum_{i=1}^{N} \frac{CF_i}{(1+r)^i} \tag{3}$$

where

CF_i means a cash flow generated in the period i.

To calculate the price p is assumed that the *NPV* must be at the time of return positive. Thus the basic assumption was that we set the maximum acceptable *payback* period of investment I in the facility F. Then n = *lifetime* = *payback* in formula (4). If we assume that

$$CF_i = pK + B_i - C_i - u_i - j_i - E_i - T_i \tag{4}$$

$$T_i = t(pK + B_i - C_i - u_i - j_i - E_i - O) \tag{5}$$

Then price p is defined as

$$p = \frac{\dfrac{I}{(1-t)\sum_{i=1}^{N} 1/(1+r)^i} - B + C + \dfrac{\sum_{i=1}^{N}(u_i + l_i + E_i)}{\sum_{i=1}^{N} 1/(1+r)^n} - \dfrac{tO}{1-t}}{K} \tag{6}$$

where

B_i is the total revenue generated from the facility in the period i;
C_i is the total operating costs arising from the facility during the period i;
K means the capacity of the facility;
T_i means a tax on income arising from the facility during the period i;
u_i means the interest due on loans for the period i;
j_i means repayment of principal on loans for the period i;
E_i means the costs of emission allowances for the period i;
i means the period (year) from 0 to n and
N means lifetime and also payback of the facility.

It is clear that different facilities will have different costs, incomes, investments, etc. For each-mentioned facilities F_j (composting, biogas plants, MBT and ERP plants, and landfills) were developed the economic submodels for the construction of the price p_j of given facility F_j (j=1…M). These models were based on the real level of investment, operating expenses, operating incomes, interest on loans, capacity of facility and emissions, Hřebíček and Soukopová [6,7,8]. The economic model of landfill was evaluated from prices for all landfills of the Czech Republic, as the average for the whole country, because the standard deviation of prices was less than 10 %.

3 Results and discussion

The above chapters briefly introduced the developed mathematical model of economics of waste treatment facilities needed for the regulation of waste management of the Czech Republic and a decision support of the allocation of subsidies from EU. This IMWM of the Czech Republic was implemented as a web-based application for evaluating the cost and price relationships for the municipal waste management of the country.

Decisions makers of the Ministry of Environment of the Czech Republic were able to use this IMWM to allocate subsidies from EU to investors of potential facilities to decline MSW from landfills to new facilities (ERP and MBT). They could choose inputs: *the list of K planned facilities* F_s ($s = 1 \ldots k$) (they are connected with their economic models); *their common payback; common value-added tax; chosen percentage of subsidy; charge of landfilling* and *landfill reclamation*. They obtained outputs of this model, where were *prices* p_s *of waste treatment* at planned *facilities* F_s, and calculated prices $CT_i = (CTF_i + CTE_i)$ for all municipalities S_i, (i=1…n) of the Czech Republic which will pay for the treatment of MSW.

4 Conclusion

The integrated waste management model of the Czech Republic has been introduced in the paper.

It enables to optimise environmental impacts. Its application was used as the decision support tool of the Ministry of Environment of the Czech Republic for optimizing EU subsidies to the planning allocation of new waste treatment facilities (ERP and MBT) with respect to expenses per capita of waste management of the Czech Republic.

The concept of the model is very general, and other additions and modifications of the model (e.g. addition of other relevant waste streams) will be performed for the future needs of its users. It can be used for modelling:

- the percentage of EU subsidy for various types of facility with respect to the total planned EU subsidies;
- the fee for landfilling and incineration;
- the number and location of facilities with regard to the quantity of MSW which is available for each facility in comparison with its planned capacity;
- the MSW treatment financial burden per capita (minimum, average and maximum) etc.

References

[1] Berger C., Savard G., and Wizere A.: EUGENE.: An optimisation model for integrated regional solid waste management planning, *International Journal of Environment and Pollution*, **12**(2/3) (1999), 280-307.

[2] Haigh M. et al.: *EPIC/CSR Integrated Waste Management Tool: Environmental Analysis Model*, Environment and Plastics Industry Council and CSR: Corporations Supporting Recycling, Version: Release 2.0.6., 2006.

[3] Hřebíček, J.: *Integrated waste management system: on regional level* (in Czech). Littera, Brno, 2009.

[4] Hřebíček, J., Hejč, M.: Quality of Data, Information and Indicators in Environmental Systems. In : *Proceedings of the 4th WSEAS International Conference on Mathematical Biology and Ecology*. WSEAS, 2008, 35–40.

[5] Hřebíček, J., and Soukopová, J.: Integrated Model of Municipal Waste Management of the Czech Republic. In *Proceedings of the Fifth Biennial Conference of the International Environmental Modelling and Software Society. Modelling for Environment's Sake*. (Swayne, D. A., Yang, W., Voinov, A. A., Rizzoli, A., and Filatova, T. eds.). International Environmental Modelling and Software Society, Guelph, Canada, 2010, 1106-1113,

[6] Hřebíček, J., and Soukopová, J.: Model of cost and price relationoships for municpal waste management of the Czech Republic, *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, **7** (2011), 371-378.

[7] Kostkan, V., Laciná, J., Mazalová, M., Hekera, P., Voženílek, V., Heisig, J.: *Aliquot structure of the mixed municipal waste in the city of Olomouc agglomeration and its seasonal dynamics during years 2008–2010*. Palacký University, Olomouc.

[8] Levy, H., Sarnat, M.: *Capital investment and financial decisions* (in Czech). Grada Publishing, Praha, 1999.

[9] McDougall, F., White, P., Franke, M., and Hindle, P.: *Integrated Solid Waste Management: a Life Cycle Inventory*, 2nd Edition. Wiley-Blackwell, New York, 2001.

[10] McDougall, F., Thomas, B. and Dryer, A.: Life Cycle Assessment for sustainable solid waste management - an introduction. *Wastes Management*, **5** (2002), 43-45.

[11] Shmelev, S. E., and Powell, J. R.: Ecological-economic modelling for strategic regional waste management systems, *Ecological Economics*, **59** (2006), 115-130.

[12] Solano, E., Ranjithan, S.R., Barlaz, M.A. and Brill, E.D.: Life-Cycle-based Solid Waste Management. I: Model Development. *Journal Environmental Engineering,* **128** (2002), 981-992.

[13] Valach, J.: *Investment decisions and long-term financing* (in Czech), 2.eddition. Ekopress, Praha, 2010.

[14] Wang F.S.: Deterministic and stochastic simulations for solid waste collection systems – A SWIM approach, *Environmental Modelling and Assessment*, **6** (2001), 249- 260.

[15] Yeomans J.S.: Application of Simulation-Optimization Methods in Environmental Policy Planning Under Uncertainty, *Environmental Informatics Archives*, **4** (2006), 167-185.

[16] Czech Act no. 185/2001 Coll. (Waste Act)
http://www.mzp.cz/C125750E003B698B/en/waste_management/$FILE/Waste%20Act_1852001.pdf.
[17] Waste Framework Directive – Directive 2008/98/EC
http://ec.europa.eu/environment/waste/framework_directive.htm.

Price competition with capacity constraint and imperfect information

Rostislav Staněk [1]

Abstract. Kreps-Sheinkman model introduces a capacity choice into the model of price competition. It shows that in the context of a one-shot game, the environment in which oligopolists make capacity decisions and then post prices for their products has the same Nash equilibrium as the environment in which oligopolists make output decisions and sell their output at market clearing prices. The aim of this paper is to investigate the robustness of this claim. I consider the two-stage game where firms make capacity decision first and then compete in prices subject to their supply limits. Contrary to Kreps and Sheinkman, I assume that firms have imperfect information about their rival's capacity choice. I show that in this case the Cournot equilibrium conclusion does not hold, because the model has no pure strategy equilibrium. Moreover, the mixed strategy equilibrium has a surprising property. Expected market price increases when there are more firms in the market. This conclusion holds for surplus-maximizing as well as proportional rationing rule.

Keywords: bertrand competition, imperefect information, capacity constraint, mixed-strategy equilibrium.

JEL classification: L10, L11
AMS classification: 91B24, 91B26

1 Introduction

In this paper, I analyze a model of the price competition with capacity choice and incomplete information. Hereby I try to reexamine some general notions regarding price competition and comparison of price and quantity competion outcomes in homogenous product market. It is well known idea that Bertrand model of price competition with constant marginal costs involves pricing at marginal costs. There are two important features of Bertrand model. First, firms compete in prices. Second, each firm is ready to supply all the forthcoming demand at the price it is setting. The first feature is considered to be an advantage of Bertrand model because it explains price determination without assuming the existence of an auctioneer. On the other hand, the second feature seems to be problematic. If we make a common assumption of decreasing returns to scale technology, then the game has multiple equilibria. Dastidar [2] shows that there may be a continuum of pure strategy equilibria, if firms must supply all the forthcoming demand once they post a price. Moreover, Hoernig [5] proves that any finite set of pure equilibrium prices that lead to positive equilibrium profits can be supported in a mixed strategy equilibrium.

There are two approaches how to incorporate the idea that a firm does not have be able or willing to supply all the forthcoming demand. First of them is to assume decreasing returns to scale technology and voluntary trading. Voluntary trading means that a firm sells at most the quantity that equals marginal costs to the announced price. Dixon [4] shows that pure strategy equilibrium in such a game exists under proportional rationing rule and quadratic costs. But in general the problem of this approach is that Nash equilibrium in pure strategies may not exist. For example, Vives [8] shows that under the surplus maximizing rationing rule the pure strategy equilibrium exists only if the competitive price is equal to the monopoly price. Another approach follows semninal paper of Kreps and Scheinkman [6] that introduces capacity choice into the model of price competition. Kreps and Sheinkman present two-stage game in which each firm chooses its capacitiy in the first stage. Capacity determines maximum possible supply of the firm. In the second stage, firms engage in Bertand price competition. They show that under

[1]Masaryk University/Faculty of Economics and Administration, Department of Economics, Lipová 41a Brno 602 00 Czech Republic, rostanek@mail.muni.cz

surplus maximizing rationing rule the subgame perfect equilibrium of the game coincides with Cournot equilibrium. The aim of this paper is to investigate the robustness of this result. The robustness of the result with respect to the rationnig rule and cost structure has been examined before. Davidson and Deneckere [3] shows that Cournot equilibrium does not emerge under proportional rationing rule. On the other hand Madden [7] claims that the result is quite robust because he shows that the Cournout output is equilibrium also under proportional rationnig rule if costs are sunk in the second stage. Unlike these papers, I change the information structure of the game by introducing incomplete infomation. I consider the same model as Kreps and Scheinkman [6] and Madden [7] with one exception. I suppose that capacity choice is not observed by other firms. I show that under incomplete information assumption the equilibrium of the model does not coincide with Cournot equilibrium. Moreover, I show that under reasonable assumptions the equilibrium exhibits very different properties from Cournot equilbrium.

2 Model

Consider a market with downward-sloping demand function $D(p)$ with finite maximum demand $D(0)$. Assume that demand function is zero for a high enough price, i.e. $D(\bar{p}) = 0$ and $\bar{p}$ represents the lowest price at which $D(p) = 0$. There are n firms in the industry. The cost functions are linear and identical, i.e. $C_i(q_i) = cq_i$. The model is structured as a two stage game. In the first stage, firms choose simultaneously their output q_i. The output determines maximum quantity that the firm is able to supply. The output decision of other firms is not observed and each firm knows only its own output decision in the second stage. In the second stage, firms choose simultaneously their prices p_i. The firm is then ready to sell at price p_i until the output q_i is reached. Each firm's strategy specifies action played at the start of the game and at each information set. Because histories (q_i, q_{-i}) and (q_i, q'_{-i}) belongs to the same information set for all q'_{-i}, each firm's strategy is a output-price pair (q_i, p_i). Preferences of each firm are given by the profit function $\Pi(p, q) = S_i(q, p)p_i - cq_i$, where $S_i(q, p)$ denotes realized sales. Consistently with Madden [7] I assume that costs are sunk in the second stage when a pricing decision is made.

Contrary to the classical Bertrand case, one firm do not have to supply all the forthcoming demand. It is therefore neccessary to specify a rationing rule for unsatisfied demand. There are two common rationing rules in the literature: surplus maximizing rationnig rule and proportional rationing rule. According to the surplus maximizing rationing rule the low-price firm serves the consumers with higher reserve prices until its capacity is reached. The surplus maximizing rationing rule would arise if consumers could engage in costless arbitrage. The realized sales of the firm under the surplus maximizing rule are given as

$$S_i(q,p) = \begin{cases} min\{q_i, D(p_i)\} & \text{if } p_i < p_j \\ min\{q_i, D(p_i)/n\} & \text{if } p_i = p_j \\ min\{q_i, max\{D(p_i) - q_j, 0\}\} & \text{if } p_i > p_j \end{cases} \tag{1}$$

The proportional rationing rule assumes that rationing at the lowest price is made through a queuing process. Hence, consumers arrive in random order to the low-price firm. The low-price firm serves arriving consumers until its capacity is reached. The total sales of the firm are then given as

$$S_i(q,p) = \begin{cases} min\{q_i, D(p_i)\} & \text{if } p_i < p_j \\ min\{q_i, D(p_i)/n\} & \text{if } p_i = p_j \\ min\{q_i, max\{(1 - \frac{q_j}{D(p_j)})D(p_i) - q_i, 0\}\} & \text{if } p_i > p_j \end{cases} \tag{2}$$

Main results of this paper hold under the surplus maximizing rationing rule as well as under the proportional rationing rule. It means that the results are directly comparable not only with the results of Kreps and Scheinkman [6], but also with results obtained by Madden [7]. Throughout the paper, the results do not depend on the empolyed rationing rule if it is not stated otherwise.

3 Equilibrium analysis

In this section, I find an equilibrium of the model and I compare the equilibrium with Cournot equilibrium obtained by Kreps and Scheinkman. First, I focus on the problem whether Cournot outcome is still an equilibrium of the model, where firms do not observe output decision of other firms. Next proposition shows that Cournot outcome cannot be an equilibrium in this case. In fact, there are no pure strategy equilibria.

Proposition 1. *There is no pure strategy Nash equilibrium in the model.*

Proof. Proof proceeds by contradiction. Suppose that the strategy profile $((q_1, p_1), ..., (q_n, p_n))$ is Nash equilibrium. I denote Φ as a set of the lowest prices p_i that satisfies the condition $q_i > 0$, i.e. $\Phi = \min\{p_1, \ldots, p_n : q_i > 0\}$. Let ϕ denotes member of this set and let q_j denotes output of firm with price ϕ. There are four possible situations $\phi < c$, $\phi = c$, $\phi > c$ and $\Phi = \varnothing$. I show that in all of these situation some firm have profitable deviation and consequently none of these situations constitutes Nash equilibrium.

1. Consider the case when $\phi > c$. If Φ is singelton and $q_j \neq D(\phi)$, then firm j can increase its profit by setting $q_j = D(\phi)$. If Φ is singelton and $q_j = D(\phi)$ or Φ has m elements where $m > 1$ then each firm can undercut its rivals and earn profit

$$\lim_{\epsilon \to 0}(\phi - \epsilon - c)D(\phi - \epsilon) > \frac{(\phi - c)D(\phi)}{m} > 0.$$

   This constitutes a profitable deviation.

2. Consider that case when $\phi = c$. If $\sum_{j=1}^{n} q_j > D(c)$, then there exist firm with positive sales and negative profit. This firm can profitably deviate by setting $q_j = 0$. If $\sum_{j=1}^{n} q_j \leq D(c)$, then there exist a firm that obtains zero profit. If this firm sets price p_j such that it is the smallest price greater than marginal costs, i.e. $p_j = \min\{p_i : p_i > c\}$, and choose some small output that is not greater than its residual demand, then it earns positive profit.

3. Consider that case when $\phi < c$. In this case firm j obtains negative profit and it can increase its profit by setting e.g. $q_j = 0$.

4. Consider the case when $\Phi = \varnothing$. This assumption implies that all firms produce zero output. Clearly, this is not a Nash equilibrium because each firm can increase its profit by setting any price $p_i \in (c, \bar{p})$ and any output $q_i \in (0, D(p_i))$.

□

Economic intuition behind this result is clear. Contrary to the Kreps and Scheinkman model, the firm cannot make a credible commitment not to produce more than the Cournout equilibrium quantity. If the firm produces greater than Cournout equilibrium quantity in the Kreps and Scheinkman model, then it is punished in the second stage because other firms set lower prices. But if the output decision is not observed, then other firms do not detect deviation from Cournot equilibrium quantity and do not set lower prices. Threfore, each firm has an incentive to produce more than Cournot equilibrium quantity.

The problem of non-existence of pure strategy equilibria can be solved by looking for mixed-strategy equilibria. The existence of mixed-strategy equilibrium will be ensured if the Dasgupta-Maskin theorem [1] can be applied. Dasgupta-Maskin theorem requires individual payoffs to be bounded and continuous, except the cases when players choose the same action, and the sum of payoffs has to be upper semi-continuous. Obviously, the sum of firm's profits is bounded and continuous in prices and output. Moreover, discontinuities in individual profits arise only when there are price ties. Hence, the assumptions of Dasgupta-Maskin theorem are met and this theorem states that mixed strategy equilibrium exists. Proposition 2 shows that there is one symmetric and a continuum of asymmetric mixed-strategy equilibria. Equilibrium strategies can be described in the following way. There are at least two firms that randomize continuously on the output interval $[0, D(c)]$ in the first stage. Other firms randomize on the output interval $[D(a), D(c)]$, where $a \leq \bar{p}$ is arbitrary price, and choose zero output with positive probability. In the second stage, each firm set such a price that it sells exactly its output supposing its price is the lowest one, i.e. $p_i = D^{-1}(q_i)$ Note that the symmetric equilibrium is a special case of the asymmetric equilibrium when all firms randomize on the price interval $[0, D(c)]$. Actually, if there are only two firms at the market, this equilibrium is unique. For the sake of calculation simplicity proposition 2 describes equilibirum stategies as a randomization over prices.

Proposition 2. *Supose that there are n firms. Let $G_i(q)$ and $F_i(p)$ denote the cumulative distribution function. The mixed-strategy weak sequential equilibrium profile* $((G_1(q_1), D^{-1}(q_1)), \ldots, (G_n(q_n), D^{-1}(q_n)))$

generates cumulative distribution function over prices $F_i(p)$ and function $q_i(p)$ given by expressions (3) and (4), where $2 \leq h \leq n$ and $a \in [c, \bar{p}]$.

$$F_i^*(p) = \begin{cases} 1 - (\frac{c}{p_i})^{1/(n-1)} & \text{if } p_i \in [c, a] \\ 1 - (\frac{c}{p_i})^{1/(h-1)} (\frac{a}{c})^{(n-h)/((n-1)(h-1))} & \text{if } p_i \in (a, \bar{p}) \quad \text{and} \quad q_i(p) = D(p), \ \forall i = 1, ..., h \\ 1 & \text{if } p_i = \bar{p} \end{cases} \tag{3}$$

$$F_j^*(p) = \begin{cases} 1 - (\frac{c}{p_j})^{1/(n-1)} & \text{if } p_j \in [c, a] \\ 1 - (\frac{c}{p_j})^{1/(n-1)} & \text{if } p_j \in (a, \bar{p}) \quad \text{and} \quad q_j(p) = D(p), \ \forall j = h+1, ..., n \\ 1 & \text{if } p_j = \bar{p} \end{cases} \tag{4}$$

Proof. There are two types of equilibrium strategies. First, I check that firm $i \in 1, \ldots, n$ do not have any profitable deviation and all actions played with positive probability bring the same expected payoff. It is clear that positive production at prices below c or above $\bar{p}$ yields negative profit. So, they can never be played in equilibrium. Expected profits are given as $\Pi_i(p_i) = (1 - F_i(p))^{h-1}(1 - F_j(p))^{n-h} p q_i(p) - c q_i(p)$. It is clear that positive production at prices below c or above $\bar{p}$ yields negative profit, so they can never be played in equilibrium. After substituting equilibrium strategies into the expected profit we get expected equilibrium profit

$$\Pi(p_i) = \left(\frac{c}{p_i}\right)^{\frac{h-1}{n-1}} \left(\frac{c}{p_i}\right)^{\frac{n-h}{n-1}} p q_i(p_i) - c q_i(p_i) = 0 \quad \text{if} \quad p_i \in [c, a] \tag{5}$$

$$\Pi(p_i) = \left(\frac{c}{p_i}\right)^{\frac{h-1}{h-1}} \left(\frac{b}{c}\right)^{\frac{n-h}{n-1}} \left(\frac{c}{b}\right)^{\frac{n-h}{n-1}} p q_i(p_i) - c q_i(p_i) = 0 \quad \text{if} \quad p_i \in (a, \bar{p}) \tag{6}$$

Hence, firm i has no profitable deviation and all actions played with positive probability give zero exected profit. Second, consider firm $j \in h+1, ..., n$. As in the previous case, positive production at prices below c or above $\bar{p}$ yields negative profit. Expected profit is given as $\Pi_j(p_j) = (1 - F_i(p_j))^h (1 - F_j(p_j))^{n-h-1} p_j q_j(p_j) - c q_j(p_j)$. Again, substituting of equilibrium strategies we get the value of expected equilibrium profit

$$\Pi_j(p_j) = \left(\frac{c}{p_j}\right)^{\frac{h-1}{n-1}} \left(\frac{c}{p_j}\right)^{\frac{n-h}{n-1}} p q_i(p_i) - c q_i(p_i) = 0 \quad \text{if} \quad p_j \in [c, a] \tag{7}$$

$$\Pi_j(p_j) = \left(\frac{c}{p_j}\right)^{\frac{h}{h-1}} \left(\frac{b}{c}\right)^{\frac{(n-h)h}{(h-1)(n-1)}} \left(\frac{c}{b}\right)^{\frac{n-h-1}{n-1}} p q_i(p_i) - c q_i(p_i) = c \left(\frac{b}{p_j}\right)^{\frac{1}{h-1}} - c < 0 \quad \text{if} \quad p_i \in (a, \bar{p}) \tag{8}$$

It is clear that also firm j has no profitable deviation and all actions played with positive probability gives zero exected profit. Therefore the strategy profile constitutes Nash equilibrium. To prove that the strategy profile is also weak sequentil equilibrium it is neccessary to show that the strategy is also sequentialy rational with respect to the firm's beliefs. Consider therefore the decision making in the second stage of the game. First, suppose the subgame after history where $q_i = 0$, then the expected profit is always zero and the firm cannot profitably deviate by choosing different action from $p_i = \bar{p}$. Second, suppose the subgame after any history where $q_i \in (D(\bar{p}), D(c)]$. Firm's belief are determined by the consistency requirement and the expected profit is therefore given by equations (5), (6), (7) and (8). It is clear that firm cannot profitably deviate by choosing different action from $p_i = D^{-1}(q_i)$. Finally, suppose subgame after any history where $q_i > D(c)$. In this case firm's belief are not determined by the consistency requirement. Suppose that firm choose p_i that maximize its expected profit with repsect to its belief. Such action is sequentialy rational and consistent with expressions (3) and (4). Hence, strategy profile consitutes weak sequential equilibrium. □

The equilibrium situation is characterized by zero expected profits despite the fact that the expected price is greater than marginal cost. This result occurs because firms compete hard enough to squeeze profits to zero. But there is also an excess production that is wasted and not consumed by the consumers. This feature is essential for every equilibrium of the model. To show that expected profit has to be zero, it is sufficient to realize that if profits are positive then the lowest price played with positive probability has to be greater than marginal cost. But if this is the case, then at least one firm can profitably deviate

by charging price $p-\epsilon$, where ϵ goes to zero. To show that expected price has to be greater than constant marginal cost one can use similar argument as the one used in the proof of propostition 1. Obviously, the equilibrium is far from being socially efficient. There are two reasons why the equilibrium is not efficient. First, the price which consumers pay for the good can be greater than the marginal costs. It can be even greater than the monopoly price and with positive probability it can be so high that no transaction takes place at all. This causes the deadweight lost. Second cause of inefficiency is the fact that some sources are wasted on production that is not consumed. It follows from the zero profit result that the value of wasted resources is equal to the expected producer's surplus.

4 Properties of the equilibrium

I have shown that the equilibria of the model are distinct from the equilibrium of the Kreps and Sheinkman model in the previous section. In this section, I analyze properties of the symmetric equilibrium presented above and I show that properties of the symmetric equilibrium are also distinct from the properties of Cournot equilibrium. In symmetric equilibrium, one firm serves the entire market demand and all consumers pay the same price. Therefore, I can define the expected market price as a price that consumers buy at. Clearly, the expected market price is $p^* = min\{p_1, ..., p_n\}$. The expected market price can be counted as individual expected price conditional on the fact that this is the lowest price times the number of firms plus price $\bar{p}$ times the probability that all firms set $\bar{p}$. Formally, the expected market price is given in the following way

$$E(p^*) = n\int_c^{\bar{p}} f(p)(1-F(p))^{n-1}\,p\,\mathrm{d}p + (1-F(\bar{p}))^n\,\bar{p} = n\int_c^{\bar{p}} f(p)c\,\mathrm{d}p + c^{\frac{n}{n-1}}\bar{p}^{\frac{-1}{n-1}}, \tag{9}$$

where $f(p) = \frac{1}{n-1}\left(\frac{c}{p^n}\right)^{\frac{1}{n-1}}$. After some algebraical manipulation the expected market price can be expressed as

$$E(p^*) = c\left(n-(n-1)\left(\frac{c}{\bar{p}}\right)^{\frac{1}{n-1}}\right) \tag{10}$$

Next propostition states how the expected market price depends on exogenous parameters of the model, namely how it depends on the number of firms and marginal costs.

Proposition 3. *The expected market price is increasing in the marginal costs. The expected market price is increasing in the number of firms.*

Proof. To prove the first part of the claim, it is sufficient to see that the derivative of expected price (10) according to the marginal cost is greater than zero.

$$\frac{\partial E(p^*)}{\partial c} = n\left(1-\left(\frac{c}{\bar{p}}\right)^{\frac{1}{n-1}}\right) > 0 \tag{11}$$

Now, I prove the second part of the claim. Taking the first derivative of expected price (10) according to the number of firms we get

$$\frac{\partial E(p^*)}{\partial n} = c\left(1+\mu^{\frac{1}{n-1}}\left(\frac{\ln\mu}{n-1}-1\right)\right), \tag{12}$$

where $\mu = \frac{c}{\bar{p}}$. If $\mu = 1$, then the derivative is zero. We know that $\mu < 1$. When we take the first derivative of the above stated expression (12) according to μ we realize that it is decreasing in the parameter μ, i.e.

$$\frac{\partial E(p^*)}{\partial n \partial \mu} = c\mu^{\frac{2-n}{n-1}}\left(\frac{1}{n-1}+\frac{\ln\mu}{n-1}-1\right) < 0. \tag{13}$$

This implies that $\frac{\partial E(p*)}{\partial n} > 0$. □

The first claim in the proposition conforms to intuition. The same predictions could have been obtained also from different models of oligopolistic competition including Cournot model or Kreps and Scheinkman model. So, the symmetric equilibrium is qualitatively similar to the Cournot equilibrium in this aspect. On the other hand, the second claim of the proposition is very counterintuitive. To explain

economic intuition behind this result, recall that firms compete sufficiently to drive down expected profit to zero. Zero profit result holds regardless of the number of firms that are present in the market. A firm has to pay the production costs everytime but its sales are positive if and only if its price is the lowest on the market. If this is not true, then afirm suffers a loss. Hence, the expected price has to be high enough to cover this loss. If there are more firms present in the market, then the probability that a firm will set the lowest price decreases and a firm suffers a loss more frequently. Therefore, the expected price has to increase. This equilibrium property is very different from the property of Cournot equilibrium. In Cournot equilibrium the market price converges to the competitive price when there is a large number of firms. This is obviously not true in model with imperfect information, because expected market price increases when the number of firms increases.

5 Conclusion

The main conclusion of this paper concerns a relationship between price and quantity competition. Kreps and Scheinkman argue that the Cournot model do not have to suppose that price is determined by an auctioneer. Cournot model can be rather interpreted as a shortcut for a two-stage game where firms choose their capacities in the first stage and engage in price competition in the second stage. The main conclusion of this paper shows limits of this interpretation. Kreps and Scheinkman result is not valid if capacity choice is not observed by other firms. Hence, it shows that a two-stage game cannot be used as a foundation of Cournot model if firms do not have perfect infomation.

The model does not seem to be very realistic. But it is important to note that the main qualitative results of the paper are still valid if only a fraction of the costs are sunk in the second stage. Therefore, the decision of firms in the first stage do not have to be interpreted as a direct production of output. It can be seen as a preparation for production which carries costs proportional to the planned production. The model then predicts that equilibria of such a situation are very bad in terms of welfare. The welfare could increase if firms were able to coordinate themselves to some correlated equlibrium.

Acknowledgements

Work on this paper has been supported by the project of specific research MUNI/A/0796/2011 of the Masaryk University.

References

[1] Dasgupta, P.–Maskin, E.: The Existence of Equilibrium in Discontinuous Economic Games II: Applications, *The Review of Economic Studies* **53** (1986), 27–41.

[2] Dastidar, K. G.: On the Existence of Pure Strategy Bertrand Equilibrium, *Economic Theory* **5** (1995), 19–32.

[3] Davidson, C.–Deneckere, R.: Long-Run Competition in Capacity, Short-Run Competition in Price, and the Cournot Model, *The RAND Journal of Economics* **17** (1986), pp. 404–415.

[4] Dixon, H.: The Competitive Outcome as the Equilibrium in an Edgeworthian Price-Quantity Model, *Economic Journal* **102** (1992), 301–309.

[5] Hoernig, S. H.: Mixed Bertrand equilibria under decreasing returns to scale:an embarrassment of riches, *Economics Letters* **74** (2003), 359–362.

[6] Kreps, D.–Scheinkman, J.: Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes, *The Bell Journal of Economics* **14** (1983), 326–337.

[7] Madden, P.: Elastic Demand, Sunk Costs and the Kreps-Scheinkman Extension of the Cournot Model, *Economic Theory* **12** (1998), 199–212.

[8] Vives, X.: Rationing rules and Bertrand-Edgeworth equilibria in large markets, *Economics Letters* **21** (1986), 113–116.

Model of the unemployment rate in the Czech Republic

Radmila Stoklasová[1]

Abstract. This paper deals with the labor market modeling in the Czech Republic. Box-Jenkins or ARIMA model is one of the approaches for modeling time series. Based in the model the proposed equation describes the unemployment rate in the Czech Republic. The unemployment rate is predicted on the basis of the model by the end of 2012. The unemployment rate is the second most watched real indicator of the state of the economy after the real income and is dealt with by most states.

Keywords: unemployment, time series, time series stationarity, autocorrelation function, partial autocorrelation function, Box-Jenkins methodology, ARIMA process, SARIMA.

JEL Classification: C32, C51
AMS Classification: 62-07

1 Introduction

Unemployment rate is one indicators of the state of the economy of the Czech Republic. The source of data of the unemployment rate is the Ministry of Labor and Social Affairs. In conclusion of the article is then by the model predicted the expected rate of unemployment in the Czech Republic by the end of 2012.

The main objectives of time series analysis are: finding patterns over time series, finding deviations from the expected course, prognosis (prediction) of the time series behavior in the future, and continuous determination of appropriate management interventions for influencing the behavior of time series. For the analysis of time series of unemployment rates here is used Microsoft Excel and statistical program SPSS (Statistical Package for the Social Sciences). For more details about this program see [7]. SPSS program allows some operations to perform time series in the standard module (Base), for a deeper analysis of time series is necessary to connect the module Trends. In terms of SPSS time series is a normal data file, and it is assumed that one row of data matrix contains observations at any one time and the rows are sorted by time so that the oldest observation is the first and the youngest observation is the last line of the matrix.

This area is important because the unemployment rate is the second most watched real indicator of the state of the economy. In the literature [3], the unemployment rate is defined as the proportion of unemployed and the labor force. The labor force is the sum of unemployed and employed. For an unemployed person is considered someone who is unemployed and actively looking. Those who don't try to get work actively are outside the labor force. Official unemployment figures underestimate actual unemployment. Sometimes, in times of recession employees shorten time jobs. This unemployment does not show in the official unemployment rate. Unemployment is often divided into frictional, structural and cyclical.

2 Registered unemployment rate in the Czech Republic - 30. 4. 2012

At 30. 4. 2012 Labor Office of the Czech Republic registered a total of 497 322 job seekers at its regional offices and their contact centers. Their number was about 27 858 less than at the end of the previous month, compared to the same period of 2011 decreased by 16 520.

During April there were 39 879 newly registered persons. Compared to the previous month it was more about 1420 persons in the same period last year about 6,800 fewer people. The records in April left a total of 67 737 applicants. It was about 12 773 more than in the previous month and 12 862 persons less than in April 2011.

Registered unemployment rate reached 8.4%. The higher unemployment rate than the national average showed 43 districts, the highest being in Bruntál (15.9%), Most (15.7%), Jesenik (15.2%), Decin (14.0%) and Hodonin (13, 9%). The lowest unemployment rate was recorded in the districts of Prague-East (3.3%), Prague Capital (4.0%), Prague-West (4.2%) and Mlada Boleslav (4.7%). Registered unemployment rate fell to men and women 7.5% to 9.7%. This article analyzes the unemployment rate for the whole Czech Republic, in [1] there are the methods for Measuring regional unemployment disparities in the Czech Republic.

[1] School of Business Administration, Silesian University, Department of Mathematical Methods in Economy, University Square 1934/3, Karviná, Czech Republic, stoklasova@opf.slu.cz

Time series analysis

The following chart shows the time course of registered unemployment rate from January 2004 to April 2012.

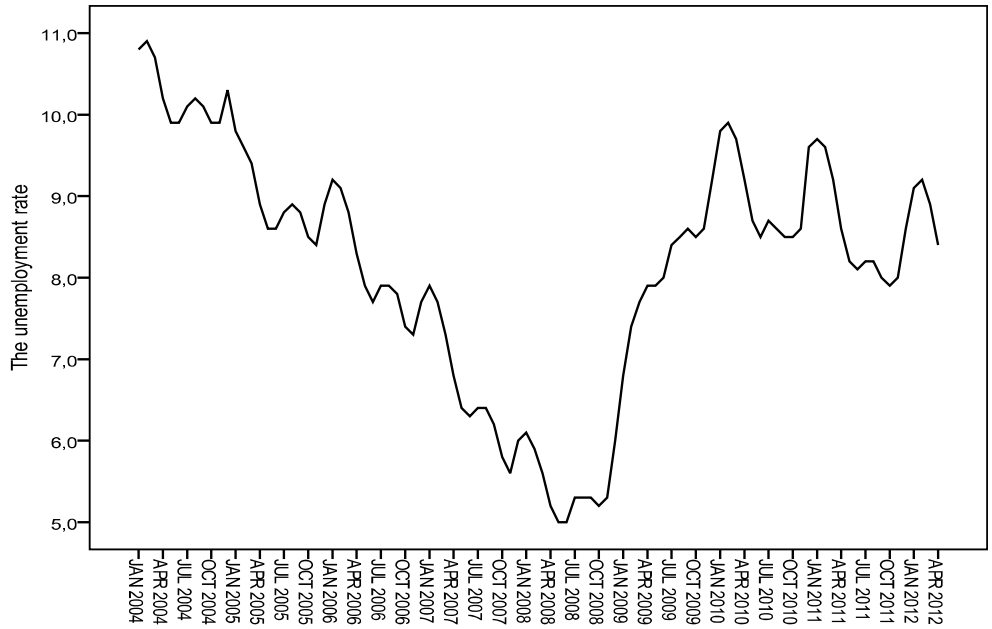


Figure 1 The registered unemployment rate

Source: Ministry of Labor and Social Affairs

The chart suggests that since February 2004 when the registered unemployment rate was highest (10.9%) during the reporting period, the time series trend downward until June 2008. Since mid-2008 to February 2010 (9.9%) had a rising trend line. The analyzed time series is nonstationary. Some methods require stationary model for the construction. Such models include the Box-Jenkins models are described in the next section. Nonstationarity of time series can be removed by an ordinary, respectively seasonal differencing time series.

3 ARIMA model

Behavior and prediction of time series is calculated by using ARIMA model. ARIMA(p, d, q) model is a complex linear model composed of three sub-sections (not always present all three): AR (Autoregressive) - linear combination of the effects of past values, I (Integrative) - random walk (designed to filter out any transient data folder), MA (Moving Average) - linear combination of past mistakes. These models are extremely flexible, they are relatively difficult to calculate and to understand the results, are demanding quality and number of measured data (assumed to be at least 50 measurements or observations). The investigated time series values, the registered unemployment rate has 100 observations, so that the ARIMA model can be used, see [2].

First, it is necessary to identify the model, i.e. determine the type of model and parameter values. This can be done based on the waveform autocorrelation (ACF) and partial autocorrelation (PACF) function of stationary series. Autocorrelation is a correlation between time series and the same time series delayed by a fixed number of measurements. Partial autocorrelation coefficients are correlation between the time series and the same series, delayed by a fixed number of measurements, the fact that the partial correlation technique eliminates the influence of intermediate members of the series. This coefficient reflects only direct link between y_t and y_{t-2} such as transmission through the exclusion of observations y_{t-1}.

In the SPSS output in the ARIMA model for each parameter there are calculated their significance. According to these values it is possible to determine whether the parameters in the model to include or not. When calculating the number of models that describe the behavior of the series about the same, then we select a model in which the AIC (Akaik‘s information criterion), respectively SBC (Schwartz-Bayesian criterion) is minimal and Log Likelihood (CAF) maximal. Finally, verify that the residual component is white noise, i.e. the sequence of random variables normally distributed with zero mean and constant standard deviation.

The ARIMA model assumes interdependence among variables $y_{t-2}, y_{t-1}, y_t, y_{t+1}, y_{t-2}, \ldots$ If the process contains even seasonal fluctuations, as in our case, we can also expect dependencies between variables in different seasons: $y_{t-2s}, y_{t-s}, y_{ts}, y_{t+s}, y_{t-2s}, \ldots$, where s is length of the period (in our case 12).

This process is called **SARIMA (*p,d,q*) (*P,D,Q*)$_s$** , where
p is order of the process AR, q is order of the process MA, d is order of the difference,
P is order of the seasonal process AR, Q is order of the seasonal process MA, D is order of the seasonal difference, s is the length of the seasonal period.
Model can be expressed by the formula:

$$\Phi_P\left(B^s\right)\phi_p(B)(1-B)^d\left(1-B^s\right)^D y_t = \theta_q(B)\Theta_Q\left(B^s\right)a_t, \tag{1}$$

where $\phi_p(B)$ is autoregressive operator, $\theta_q(B)$ is the operator of moving averages, $\Phi_P\left(B^s\right)$ is seasonal autoregressive operator, $\Theta_Q\left(B^s\right)$ is seasonal operator of moving averages, $\{a_t\}$ is white noise.

3.1 Identification of the model

First, the parameters of the model are estimated using autocorrelation and partial autocorrelation functions. The time series is nonstationary, therefore it must be differentiated by the first order. The time series exhibits seasonality, that is why it is also seasonally differentiated.

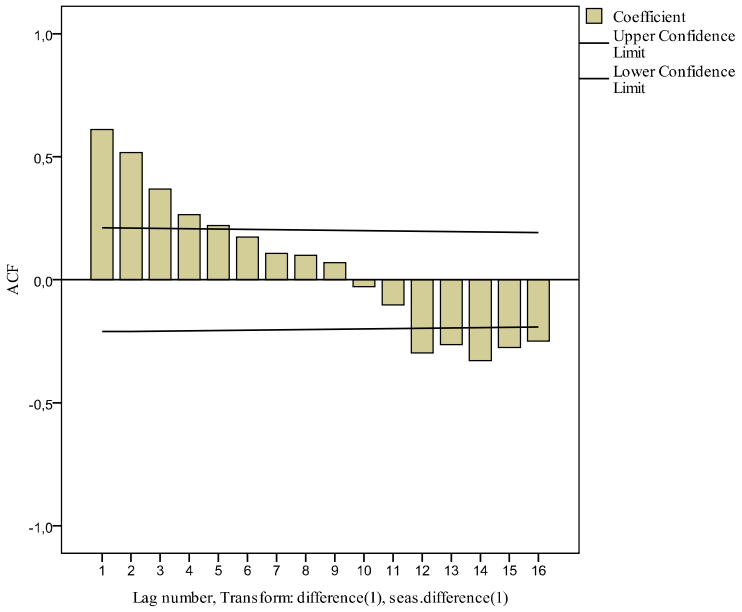


Figure 2 Autocorrelation

Based on the functions we can chose models: SARIMA (1,1,0) $(1,1,0)_{12}$ or model SARIMA (1,1,0) $(0,1,0)_{12}$. Because Akaik's information criterion as well as Schwartz-Bayes criterion are less for the model SARIMA (1,1,0) $(1,1,0)_{12}$, we chose that model.

SPSS output for model SARIMA (1,1,0) $(1,1,0)_{12}$

FINAL PARAMETERS:

Log likelihood	23.013138
AIC	-42.026277
SBC	-37.094461

Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
AR1	0.66163944	0.08297210	7.9742404	0.00000000
SAR1	-0.47709376	0.10673931	-4.4697098	0.00002402

General form of the model is: $$\Phi_1\left(B^{12}\right)\phi_1(B)(1-B)^1\left(1-B^{12}\right)^1 y_t = a_t\,. \tag{2}$$

After adjustment equation and substituting the estimated values into (2) we obtain the following equation, which describes the dynamics of the investigated time series:

$$\begin{aligned} y_t = {} & 1.66\,y_{t-1} - 0.66\,y_{t-2} + 0.53\,y_{t-12} - 0.88\,y_{t-13} + 0.35\,y_{t-14} + \\ & + 0.47\,y_{t-24} - 0.78\,y_{t-25} + 0.13\,y_{t-26} + a_t \end{aligned} \tag{3}$$

It can be inferred from the model equation, the registered unemployment rate in the current month is most affected by the value rates in the previous month and the value 13 months ago.

3.2 Diagnostic model control

Basic empirical diagnostic check is to assess residues $$\hat{a}_t = \left[\hat{\theta}_q(B)\right]^{-1}\hat{\phi}_p(B)y_t\,, \tag{4}$$

where $\hat{\phi}_p(B) = 1 - \hat{\phi}_1 B - ... - \hat{\phi}_p B^p$ and $\hat{\theta}_q(B) = 1 - \hat{\theta}_1 B - ... - \hat{\theta}_q B^q$.

Autocorrelation of unsystematic components can be tested using the sample autocorrelation fiction

$$r_k = \frac{\sum_t \hat{a}_t \hat{a}_{t-k}}{\sum_t \hat{a}_t^2}\,. \tag{5}$$

If unsystematic components are not autocorrelated, the value of this function should lie within the range $\pm 2/\sqrt{n}$ (95% confidence interval), where $n = 100$.

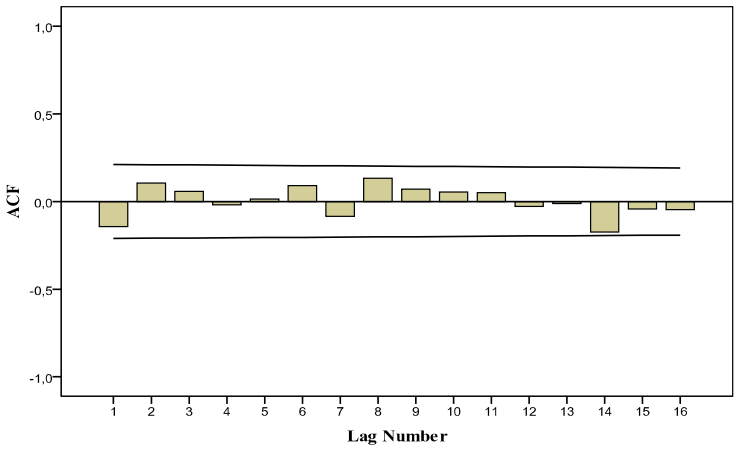


Figure 3 Autocorrelation of residues

Another way to determine whether non-systematic component is autocorrelated, is using of portmanteau test, which suggested Box and Pierce.

There is tested hypothesis H_0: $\rho_1 = \rho_2 = ... = \rho_K = 0$ against hypothesis H_1: non H_0, where ρ_k, $k = 1,..., K$ is autocorrelation of unsystematic model components for the delay k. If ARIMA model is properly constructed, then the statistics $$Q = n\sum_{k=1}^{K} \hat{r}_k^2\,, \tag{6}$$

for high n and K has approximately distribution χ^2 with $(K - p - q)$ degrees of freedom.

Lag	Autocorrelation	Lag	Autocorrelation
1	– 0.143	9	0.071
2	0.106	10	0.055
3	0.057	11	0.05
4	–0.019	12	–0.029
5	0.013	13	–0.012
6	0.091	14	–0.174
7	–0.085	15	–0.043
8	0.132	16	–0.046

Table 1 Autocorrelation of residues

The statistics $Q = 11.41$ and $\chi^2_{0,05}(15) = 24.9$. The hypothesis H_0 can not be reject. There was no evidence of correlation of residues.

3.3 Short-time prediction by estimated model

The course of the original time series of registered unemployment rate and the time series, which was calculated by the above SARIMA model, is presented in the following graph.

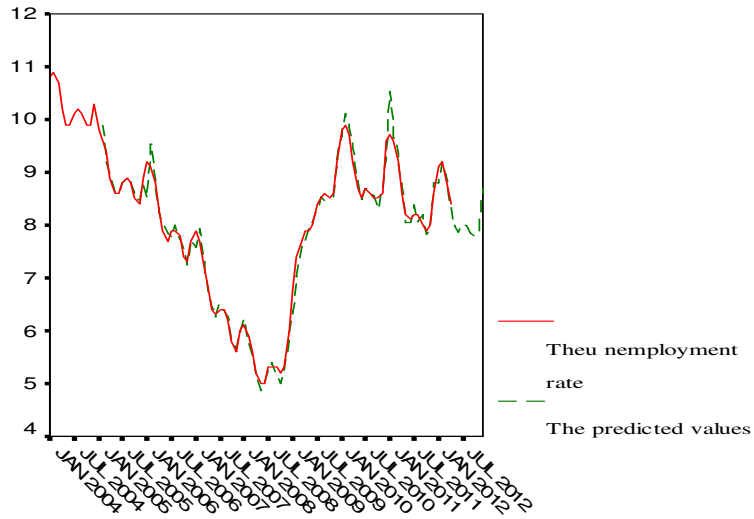


Figure 4 Graph the real and predicted time series

Source: Ministry of Labor and Social Affairs and own calculations

Point and interval forecasts (95%) of the registered unemployment rate are shown in Table 1.

Month	Point prediction	Interval prediction
May	7.99	(7.6 ; 8.4)
June	7.86	(7.2 ; 8.6)
July	8.02	(7.0 ; 9.1)
August	7.99	(6.6 ; 9.3)
September	7.84	(6.2 ; 9.5)
October	7.79	(5.9 ; 9.7)
November	7.90	(5.7 ;10.1)
December	8.69	(6.3 ;11.1)

Table 2 Prediction of the unemployment rate in the Czech Republic

Source: own calculations

4 Conclusions

In 2012 we can expect the highest values of the registered unemployment rate during the holiday period, which may be caused by registration of graduates of secondary schools and universities to the labor offices. The lowest value of the unemployment rate is expected in October 2012 (7.79%). The reason may be seasonal work in construction, as shown for example in [6].

With unemployment facing many countries, therefore it is important to capture the trend in unemployment. ARIMA model used is a highly flexible, so it is highly likely, that unemployment will develop according to the above model if there is absence of government intervention that would significantly change this trend.

References

[1] Antoušková, M.: *Measuring regional unemployment disparities in the Czech Republic.* Ostrava: Vysoká škola báňská, Území, znalosti a rozvoj na počátku 21. století: sborník příspěvků odborné sekce z konference Zvyšování konkurenceschopnosti, aneb, Nové výzvy pro rozvoj regionů, států a mezinárodních trhů, 2007, s. 153-160.

[2] Arlt, J., Arltová, M.: *Ekonomické časové řady*. Praha: Professional Publishing, 2009.

[3] Dornbusch, R., Fischer, S.: *Makroekonomie.* Praha: SPN, 1994.

[4] Mielcová, E.: *Unemployment rate and new cars registration - the case of the Czech republic 2003-2009.* Opava: Slezská univerzita, Obchodně podnikatelská fakulta v Karviné: Acta academica karviniensia, 2010, s. 348-353.

[5] Ministerstvo práce a sociálních věcí. *Informace o počtu nezaměstnaných v České republice*, duben 2012. http://portal.mpsv.cz/sz/stat/nz/mes

[6] Ministry of Labour and Social Affairs. *Basic indicators of labour and social protection in the Czech Republic : time series and graphs 2007*. Prague: Ministry of Labour and Social Affairs, 2008.

[7] Norušic, M. J. : *SPSS Professional Statistics 7.5.* Chicago: SPSS Inc., 1997.

Crop production function - study

Tereza Suchánková[1], Radka Bezděkovská[2]

Abstract. In general, the agricultural crops are significantly dependent on climate factors. Such variables could be not only the temperature and rainfall but also the soil moisture and the level of evaporation. Nevertheless the other factors have also an important role in the variability of crop production - classical production factors (capital, land and labor) or fertilizers and pesticides. The modeling of crop production is complex problem and needs the sophisticated approach. The aim of the paper is to study the most appropriate form of crop production function - the applied variables and mathematical form. The CES and VES production function would be the convenient tool. The estimation is based on panel data (time and Czech districts/regions) and on the outputs of hydro-meteorological model e-Hype. Application of panel data enables to capture the regional differences. The results of the analysis include the comparison of the sensibility of different crops on climate factors in regional level.

Keywords: production function, crop yield, planted area, nonlinear VAR model, panel data.

JEL Classification: C23, C51
AMS Classification: 91B38, 91B76

1 Introduction

The analysis of agricultural production and its planning can be studied from two different points of view. The first one consists in analyzing the crop yield as a function of weather factors, the amount of used pesticides and fertilizers and the classic production factors (labor and capital). The monthly rainfall and temperature (or other climate variables like soil moisture and evaporation) could be highly significant in explaining yield but at the same time these variables are hardly predictable. The first approach helps in climate influence analysis – how much the different crops are climate sensitive in comparison with used technology of production.

The second point of view is oriented on farmers` decision-making. The main interest of farmer is to decide how many acres would be appropriate to seed. The surface of planted area could be explained by lagged variables: planted area (autoregressive model), Agricultural Producers Price Index (APPI), Consumer Price Index (CPI) and direct costs on crop production (labor, pesticides, fertilizers, fuels…). The planted area prediction model has a character of autoregressive model – the stationarity of time series must be tested and the most convenient mathematical form of model as well.

Concerning the climate sensitivity model, the influence of weather factors in the region (mainly represented by average monthly rainfall and temperature) on the real harvest will be tested. The importance of specific climate factors, i.e. evaporation and soil moisture, will be evaluated also in relation with rainfall and temperature. The crops are weather sensitive in diverse manner depending on their biological nature which is strongly related with their production function – not only the most appropriate set of variables but also the character of elasticity of substitution (unit, constant or variable).

The various mathematical forms of mentioned models will be tested in terms of their degree of conformity with reality and of course the statistical significance of the estimated parameters. Using existing panel data (Czech regions from 2002 to 2011) we will include the influence of individual differences of each region in the model, either in form of fixed or random effects.

2 Theoretical background

Production function allows to explain the output value generated by either company, industry or the whole economy based on diverse combinations of factors determining the existing technology. Detailed explanations of the Cobb-Douglas production function (PFCD) used in the analysis can be found in [4]. The same literature explains the principles of the constant elasticity of substitution production function (PFCES), including well-known

[1] University of Economics Prague, Department of Econometrics, W. Churchill Sq. 4, 13067 Prague 3, Czech
[2] Sysnet s.r.o., Kroftova 1, 15000 Prague 5, Czech Republic, bezdekovska@gmail.com.

Kmenta's approximations for two inputs. Detailed description can be found in [5]. The variable elasticity of substitution production function (PFVES) can be found in [7].

The second type of model, i.e. vector autoregression model (VAR model), is presented in [3] and the analysis of stationarity of time series is also in [2]. The farm supply theory is analyzed in [1]. Theoretical background and estimation methods for working with panel data can be found for example in [3] or [6] – fixed and random effects are sufficiently detailed. The production functions mentioned above – PFCD, PFCES and PFVES are described below.

2.1 Cobb-Douglas production function (PFCD)

The following equation shows the well-known PFCD with two production factors:

$$Y_{it} = a_i e^{gt} A_{it}^{\alpha} B_{it}^{\beta} e^{u_{it}}, \tag{1}$$

where Y_{it} is a crop yield in the region i and in the time t, a_i represents the level of achieved technology in the region i, g the non-objectified technological progress – the parameter for the proxy variable time t, A_{it} and B_{it} are the explanatory variables – production factors in the region i and in the time t. The coefficients α and β are the elasticities of output (harvest) with respect to the explanatory variables. The sum of these elasticities gives the information about the returns to scale. The u_{it} is the stochastic disturbance term. For more information about the individual effects of each region, it is useful to log-transform the model in order to obtain the form linearized in parameters. Then, we have:

$$\ln Y_{it} = \ln a_i + gt + \alpha \ln A_{it} + \beta \ln B_{it} + u_{it}. \tag{2}$$

2.2 Constant elasticity of substitution production function (PFCES)

CES production function is more general than PFCD. The elasticity of substitution may not be the unit, only needs to be constant. This function assumes returns to scale not necessarily equal to one. To capture the individual effects it is necessary to use the linearized form of the function. Given the issue and the number of relevant explanatory variables CES function for two inputs is very restrictive, but we will try to demonstrate the function application suitability using a limited set of inputs. The CES function has the following form:

$$Y = e^{gt} c \left[\gamma A^{-\rho} + (1-\gamma) B^{-\rho}\right]^{-r/\rho} e^{u}, \tag{3}$$

where e^{gt} is non-objectified technological progress, c is the parameter of efficiency of the production process, γ is the distribution parameter depending on the units of both factors A and B, r is the degree of homogeneity and ρ is the substitution parameter. The parameters can be estimated by nonlinear least squares.

As mentioned above, to incorporate the fixed or random individual effects, the linear form of the function is needed. Linear approximation is given by Kmenta in [5]. The model with individual effects has form:

$$\ln Y_{it} = \alpha_i + gt + \beta_1 + \beta_2 \ln A_{it} + \beta_3 \ln B_{it} + \beta_4 \left[\ln\left(\frac{A_{it}}{B_{it}}\right)\right]^2 + u, \tag{4}$$

where α_i is a deviation from the constant, representing the influence of each individual region i, β_1 is common constant, g the non-objectified technological progress – the parameter for the proxy variable time t, β_2 and β_3 are the elasticity coefficients of the explanatory variables A and B and β_4 expresses the elasticity of correction part of the model. The compliance of the random component with Gauss-Markov assumptions is expected. Using the estimated parameters β the estimates of the initial parameters of PFCES can be obtained:

$$c = e^{\beta_1}, \quad r = \beta_2 + \beta_3, \quad \gamma = \frac{\beta_2}{\beta_2 + \beta_3}, \quad \rho = \frac{-2\beta_4}{\beta_2 \beta_3}\left(\beta_2 + \beta_3\right). \tag{5}$$

2.3 Variable elasticity of substitution production function (PFVES)

CES production function requires the constant elasticity of substitution in all points of an isoquant sited in isoquant map. The PFVES relaxes this requirement and supposes that the elasticity of substitution is constant only along a ray drawn from zero through the isoquant map but the substitution parameter can vary along an isoquant. The VES production function has the form:

$$Y = \gamma K^{\alpha(1-\delta\rho)}\left[L + (\rho - 1)K\right]^{\alpha\delta\rho},$$
$$\gamma > 0, \quad \alpha > 0, \quad 0 < \delta < 1, \quad 0 \le \delta\rho \le 1, \quad \frac{L}{K} > \left(\frac{1-\rho}{1-\delta\rho}\right), \tag{6}$$

where Y is output, K is capital, L is labor and α, γ, δ and ρ are parameters. Detailed properties of PFVES are described in [7].

3 Model formulation and estimation

3.1 Crop weather sensitivity analysis

Regarding the aim of the article the crop production function model assumes that dependent variable, i.e. the crop yield in metric tons, is affected by different types of climate explanatory variables – the average monthly values of temperature, rainfall, evaporation and soil moisture. To estimate the production function we utilize the available annual panel data reflecting the differences in the various regions of the Czech Republic for the period between 2002 and 2011.

Table 1 shows the correlation coefficients that represent the values for the one selected region (Central Bohemia region) to illustrate the various impacts of explanatory variables on crop yield (in this case wheat). The explanatory variables in the table below correspond to the factors mentioned above (average monthly value for T3-T7 – temperature from March to July, R3-R7 – rainfall from March to July, the period between wheat sprouting and harvesting, then E6-E7 and SM6-SM7 for the evaporation and the soil moisture in June and July).

T3	T4	T5	T6	T7	R3	R4	R5	R6	R7	SM6	SM7	E6	E7
-0.24	-0.05	-0.58	-0.72	-0.40	0.14	0.29	-0.01	0.44	0.11	0.35	0.29	0.50	0.01

Table 1 Correlation coefficients between crop yield (wheat) and climate variables in Central Bohemia region

The values of correlation coefficients presented in Table 1 indicate the possibility of existence of some dependency between the crop yield and weather factors. The following subsections will focus on estimating the different crop production function based on assumptions of Cobb – Douglas, CES and VES production functions. In the following models we assume the influence of non-objectified technological progress that is expressed through a proxy variable time. The econometric software Gretl was used to estimate the parameters and statistical characteristics.

The influence of the aforementioned explanatory variables on **wheat yield** was first expressed using the PFCD. To capture the individual effects caused by the different nature of each region the form of fixed and then random effects was used. As stated above, we assume that the wheat harvest is influenced by climate factors such as average monthly temperature, rainfall, evaporation and soil moisture. Given the number of these variables and given the length of time series and the number of regions, if we would include them all at once into the model, the estimated parameter of variables that really have an important impact on the dependent variable would have not have too significant t-test or may be even insignificant. For this reason, we will examine the impact of these variables separately. The results from this partial analysis would determinate which variables are important to explain the variability of wheat yield and which are not.

Table 2 shows the OLS estimation results for temperature variables and the values reveal that all estimated parameters are significantly different from zero at 1 % confidence level except the variable designing the average monthly temperature in March that is not significant at all. On the other hand, all coefficients do not satisfy the basic assumptions of PFCD. Their values don't lie in the interval (0, 1) – they are below zero so they affect the wheat harvest negatively. Coefficient of determination and its adjusted version is very high and close to one.

Variable	Coefficient	t-Statistic	Prob.	Variable	Coefficient	t-Statistic	Prob.
t	0.023705	4.1576	0.0001***	$\ln(a_i)$	-27.8657	-2.3919	0.0183**
$\ln(T3_{it})$	-0.023798	-0.7429	0.4590	$\ln(T6_{it})$	-0.68832	-3.4059	0.0009***
$\ln(T4_{it})$	-0.34205	-4.2013	0.0001***	$\ln(T7_{it})$	-0.70539	-11.0131	0.0000***
$\ln(T5_{it})$	-0.96309	-7.3925	0.0000***				
R-squared		0.984964		Adjusted R-squared		0.982583	

Table 2 Wheat – OLS estimation results – PFCD, fixed effects – explanatory variables *t*, T3-T7

The results from estimation for the PFCD with average monthly rainfall variables are similar; there is no need to state them. The evaporation and soil moisture variables were not statistically significant so we assume that those variables don't explain well the wheat yield.

As shown above, the selected production and climatic factors well explain the fluctuations in the harvest. Unfortunately, we can also conclude that the use of the PFCD to estimate wheat harvest is inappropriate, almost all of the explanatory variables violate basic assumptions of this production function. All this implies that *wheat production function has not the character of unit elasticity of substitution*. This signifies that PFCES may have the convenient attributes for estimating wheat harvest. Table 3 presents the results of OLS estimation of PFCES – the set of explanatory variables that have the most significant influence.

Variable	Coefficient	t-Statistic	Prob.	Variable	Coefficient	t-Statistic	Prob.
t	0.02402	8.6467	0.0000***	β_1	-45.3819	-8.1041	0.0000***
$\ln(T6_{it})$	-0.56791	-5.5200	0.0000***	$\ln(T6_{it}/$	-0.03331	-2.4684	0.0150**
$\ln(R7_{it})$	0.11564	3.2364	0.0016***	$R7_{it})^2$			
R-squared		0.587745		Adjusted R-squared		0.530299	

Table 3 Wheat – OLS estimation results – PFCES, fixed effects – explanatory variables *t*, T6 and R7

All estimated parameters are statistically significant at 1% confidence level except β_4 (5% confidence level). The coefficient β_4 (-0.03331) is statistically significant - if not, it would be more appropriate to use PFCD. The parameter ρ is 0.459, its positive value indicates the elasticity of substitution less than 1 and induces that translog approximation may be used. The wheat yield is explained in 59% by the model. The parameter β_2 represents the negative influence of higher temperatures in June. Contrary, the rainfall in July has a positive effect. The value of F-test (7.2351) with p-value = 2.559e-10 demonstrates the existence of important differences between regions - the model with fixed individual effects is convenient - Hausman test proved that in random effects model the parameters estimated by GLS aren't consistent. The Akaike information criterion (AIC=-253.6) has the best value from all tested models. All the results mentioned above show on *the appropriateness of the CES production function to model the wheat yield*. We apply this analyzing process to the other crops: rye, barley and corn.

Table 4 denotes results of OLS estimation of PFCD for **rye yield**– the set of explanatory variables with most significant parameters.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
$\ln(R7_{it})$	0.25016	0.11914	2.0997	0.03778**
$\ln(T3_{it})$	0.358195	0.114741	3.1218	0.00224***
β_1 – constant	6.97798	0.611034	11.4199	0.00000***
R-squared	0.809870	Adjusted R-squared		0.786870

Table 4 Rye – OLS estimation results – PFCD, fixed effects – explanatory variables T3 and R7

The estimation shows that all parameters are statistically significant. The value of F-test (38.375) with p-value = 9.88e-38 demonstrates the existence of important differences between regions. Testing the form of individual effects, Hausman test didn't reject the hypothesis that GLS estimates are consistent (for random effects). We used the Akaike information criterion to decide between two models – fixed effects (AIC=314.9) and random effects (AIC=519.1). The fixed effects are more appropriate. *All estimated parameters are the values complying the assumptions of PFCD function.* The CES and VES production functions didn't give significant results.

The results of OLS estimation of PFCD for **barley yield** are shown in Table 5.

Variable	Coefficient	t-Statistic	Prob.	Variable	Coefficient	t-Statistic	Prob.
t	0.01362	3.2517	0.00148***	β_1	37.317	4.4776	0.00002***
$\ln(T4_{it})$	0.39928	8.7101	0.00000***	$\ln(R7_{it})$	0.08001	3.2478	0.00150***
$\mathrm{Ln}(R3_{it})$	0.043014	2.1639	0.03242**				
R-squared		0.988108		Adjusted R-squared		0.986451	

Table 5 Barley – OLS estimation results – PFCD, fixed effects – explanatory variables *t*, T4, R3 and R7

Table 5 shows statistical significance of all parameters. The presented estimation considered only the set of the most convenient variables. The value of F-test (756.18) with p-value=7.06e-110 demonstrates the existence of important differences between regions. The fixed individual effects showed better results (AICfixed=-185.8, AICrandom=405.8) even if Hausman test didn't reject the null hypothesis that GLS estimates in random effects model are consistent. All estimated parameters are in needed interval so the PFCD well describes the barley yield. In addition to this fact, the estimated parameters of PFCES and PFVES weren't significant.

Table 6 summarizes the results of OLS estimation for **corn yield**. As shown below, the production function of dependent variable has constant elasticity of substitution – the PFCES well describes the behavior of corn yield. Neither PFCD nor PFVES estimates were statistically significant.

Variable	Coeff.	t-Statistic	Prob.	Variable	Coeff.	t-Statistic	Prob.
β_1	14.6932	20.9719	0.00000***	$\ln(R3_{it})$	0.06478	1.8019	0.07400*
$\ln(T7_{it})$	-0.81162	-3.3196	0.00119***	$\ln(T7_{it}/\ R3_{it})^2$	-0.08123	-2.8572	0.0050***
R-squared		0.978812		Adjusted R-squared		0.976056	

Table 6 Corn – OLS estimation results – PFCES, fixed effects – explanatory variables T7 and R3

The high statistical significance of estimated parameters, mainly the coefficient β_4, show the appropriateness of PFCES. The corn yield is explained almost in 98%. The value of F-test (417.02) with p-value = 5.4e-95 point on the existence of important differences between regions. The type of individual effects may be evaluated by Hausman test but the test statistic can't reject the null hypothesis about consistency of GLS estimates. Random effects may be used as well as fixed. Then, we can decide between fixed and random effects using the Akaike information criterion (AICfixed = -7.02 and AICrandom = 505.59) – fixed effects are more convenient tool.

Table 7 demonstrates the **differences in weather sensitivity of different crops**. Every crop is characterized by the most important weather factors and related sensitivity coefficients that represent the intensity of influence and the direction – negative or positive.

Crop	Elasticity of substitution	R-squared	Weather factor	Sensitivity
Wheat	Constant	0.588	T6	-0.57
			R7	0.12
Rye	Unit	0.810	T3	0.36
			R7	0.25
Barley	Unit	0.988	T4	0.40
			R3	0.04
			R7	0.08
Corn	Constant	0.979	T7	-0.81
			R3	0.06

Table 7 Comparison of crop weather sensitivity

3.2 Planted area prediction model

The planted area (A) in time *t* could be explained by lagged variables (in time *t* -1): planted area (autoregressive model), Agricultural Producers Price Index (APPI), Consumer Price Index (CPI) and direct costs (DC) on crop production (labor, pesticides, fertilizers, fuel…). The estimation method for VAR model presumes the stationarity of time series. The non-stationarity may be eliminated by including time variable or first differences of dependent variable. The planted area doesn't evince any visible trend. We applied the unit root test (Augmented

Dickey-Fuller test) that rejected the null hypothesis that all groups have unit root. The VAR model form of linear regression and power-law function applied on panel data of wheat crop was tested. The non-linear function (Cobb-Douglas type) showed better results. The linearized model has a form:

$$\ln A_{it} = gt + c_1 + c_2 \ln A_{i,t-1} + c_3 \ln APPI^1_{i,t-1} + c_4 \ln APPI^2_{i,t-1} + c_5 \ln CPI_{i,t-1} + c_6 \ln DC_{i,t-1} + u_{it}, \quad (7)$$

where t is time variable and g its parameter, c_1 is a constant and parameters c_2 to c_6 correspond to the influence of lagged dependent (A) and explanatory variables: price index of a studied crop - wheat $APPI^1$, price index of a competing crop – rye $APPI^2$, CPI and direct costs DC. Table 8 presents the results of OLS estimation.

Variable	Coefficient	t-Statistic	Prob.	Variable	Coefficient	t-Statistic	Prob.
g	0.02046	8.2220	0.00000***	c_4	-0.12009	-4.3238	0.00003***
c_1	-27.9551	4.7445	0.00000***	c_5	1.14872	5.4983	0.00000***
c_2	-0.18314	-2.9331	0.00411***	c_6	-0.47630	-7.6641	0.00000***
c_3	-0.13410	-4.8649	0.00000***				
R-squared		0.996628		Adjusted R-squared		0.996024	

Table 8 Wheat – OLS estimation results – Planted area prediction model, fixed effects

As shown above, the estimated parameters are all statistically significant at 1% confidence level. The model well describes the variability of dependent variable. The high value of determination coefficient R-squared may denote the spurious regression. Granger and Newbold [2] show that the relation R-squared > DW statistic may signify that the residuals have non-stationary character. In that case, DW statistic is 1.202 and mentioned hypothesis could be rejected. In evaluation of the importance of regional individual effects, the F – test (F = 27.64, p-value = 3.22e-28) rejected the null hypothesis that the groups have common intercept – the individual effects of different regions have important influence. The model with random effects doesn't explain the variability of dependent variable because the Hausman test rejected the null hypothesis that GLS estimates are consistent. The Akaike information criterion (AIC = -330.26) for the fixed effects model has the best value in comparison with other tested forms of model. Due to the use of panel data, low value of DW statistic does not necessarily mean that the autocorrelation is present.

4 Conclusion

The paper presents the analysis of crop weather sensitivity and planted area prediction model. The first part of paper demonstrates that every crop needs to be considered separately. Wheat and corn could be represented by PFCES but the production of rye and barley has character of unit elasticity of substitution – PFCD. Every analyzed crop was influenced by different weather factors and their importance varies – due to the crop biological nature. The second part of the paper concentrates on planted area predictions. The tests show that the model has autoregressive character and has the form of power-law function. The important influence of lagged dependent and explanatory variables (agricultural producers price index for the studied crop and for the competing crop, consumer price index and direct cost of production) is proved.

Acknowledgements

This work was supported by Internal Grant Agency of the University of Economics in Prague - IGA F4/18/2012 and project SUDPLAN (European Union).

References

[1] Ferris, J.N.: *Agricultural prices and commodity market analysis*. Michigan State University Press, East Lansing, Michigan, 2005.
[2] Granger, C.W., J.-Newbold, P.: Spurious Regression in Econometrics, *Journal of Econometrics* **2** (1974), 111-120.
[3] Greene, W.H.: *Econometric analysis, 5th ed.* Prentice Hall, Upper Saddle River, New York, 2003.
[4] Hušek, R.: *Aplikovaná ekonometrie.* Vysoká škola ekonomická v Praze, Praha, 2001.
[5] Kmenta, J.: On estimation of the CES production Function. *International Economic Review* **8** (1967),80-89.
[6] Koop, G.: *Introduction to Econometrics.* John Wiley & Sons Ltd., England, 2008.
[7] Revankar, N.S.: A Class of Variable Elasticity of Substitution Production Functions. *Econometrica* **39** (1971), 61-71.

Application of Markov chain analysis to trend prediction of stock indices

Milan Svoboda [1], Ladislav Lukáš [2]

Abstract. The paper concerns with study aims at trying to predict the stock index trend of Prague stock exchange PX using Markov chain analysis (MCA). The prediction of the trend using MCA is done using time series of day closing prices from Jan.5, 2004, till Dec.12, 2009. Downloaded data are processed in order to provide current price rates and simple technical indicator, as well. Discrete state spaces are defined for four MCA models, and appropriate transition probability matrices are calculated. These objects represent a core of any MCA and its application. The results of the short-term trend prediction using MCA are reported for various investment strategies. Numerical calculations and computer implementations have been done by MS-Excel and Mathematica modules which are briefly discussed as well.

Keywords: Markov chain analysis, transition probability matrix, stock index, trend prediction, time series analysis.

JEL Classification: C02, C13, G14, G19
AMS Classification: 90C40, 91B82

1 Introduction

The prediction of financial market is a complex task since the distribution of financial time series is changing over a period of time. There is also never ending debate as to whether these markets are predictable or not. In other words, they are called efficient markets (EMH) if being unpredictable ones, and vice versa. In the recent years, investors have started to show interest in trading on stock markets indices as it provides an opportunity to hedge their market risk, and at the same time if offers a good investment opportunity for speculators and arbitrageurs.

There is a dream of fascination of any investor to know the future asset prices and/or any financial instrument, e.g. stock exchange index. Basically, traders use several different approaches for prediction based upon fundamental analysis, technical analysis (TA), psychological analysis, etc. The technical analysis paradigm states that all price relevant information is contained in market price itself. Hence, the instant processing of market messages plays specific role, thus leading to permanent interactions among traders. TA concerns with identifications of both trends and trend reverses using more or less sophisticated procedures to predict future price movements from those of the recent past.

There is well-known that predictability of prices is a matter of research, discussion and accumulation of empirical evidence with lot of both for and against papers and works. Academicians, who belief on the EMH, see Fama [3], have been rather sceptic thereabout, but a lot of skilled practical traders still belief in opposite. Further, there are available other works bringing critical analysis of TA as well, see Fama and Blume [4], and Jensen and Benington [5]. However, there are also some later studies which show TA methods to be capable of outperforming the market, see Sweeney [6], and Brock at al. [1], as an example. Hence predicting stock market trend has become an important activity. Finally, still increasing computer power and the development of large financial oriented databases both accelerate number of recent works and studies focused upon TA and their various methods with the main goal to analyse their profitability in probabilistic context.

2 Markov chain models

The present study aims at trying to predict stock market asset prices using Markov chain models. In words, a Markov chain (MC) is a special kind of stochastic process where the next state of the system depends only on the current state and not on the previous ones. Stochastic process in form of discrete sequence of random variables $\{X_n\}$, n=1,2,.. , is said to have the Markov property if (1) holds for any finite n, where particular realizations x_n belong to discrete state space $S = \{s_i\}$, i=1,2,..,k. Generally, MC is described by vectors $\boldsymbol{p}(n)$ which give unconditional probability distributions of states, and transition probability matrix $\boldsymbol{P}$ which gives conditional probabilities $p_{ij} = P(X_{n+1} = s_j \mid X_n = s_i)$, i,j=1,2,..,k where p_{ij} may depend on n. In such case we speak about non-homogeneous

[1] University of West Bohemia, Faculty of Economics, Husova 11, 30614 Plzeň, svobodm@kem.zcu.cz .
[2] University of West Bohemia, Faculty of Economics, Husova 11, 30614 Plzeň, lukasl@kem.zcu.cz.

MC on the contrary to homogeneous MC where p_{ij} does not depend on n at all. Development of $\boldsymbol{p}(n)$ is given by recurrence equation (2), where $^{\mathrm{T}}$ denotes transposition.

$$P\left(X_{n+1} = x_{n+1} \mid X_1 = x_1, X_2 = x_2, .., X_n = x_n\right) = P\left(X_{n+1} = x_{n+1} \mid X_n = x_n\right) \tag{1}$$

$$\boldsymbol{p}(n+1)^{\mathrm{T}} = \boldsymbol{p}(n)^{\mathrm{T}}\, \boldsymbol{P},\ n=1,2,.. \tag{2}$$

The application of MC to stock market analysis is not new. From the recent works we refer to Doubleday and Esunge [2], and Vasanthi et al. [7]. In our study we concern modeling of Prague stock exchange index PX and its trend development by MC. We have at our disposal the set of PX day closing prices P_t from Jan.1, 2004, till Dec.12, 2009, and we calculate two time series Y_t and K_t thereof, which serve us further to define various state spaces S and corresponding MC models, too. Y_t is a chain index of day closing prices given by $Y_t = P_t/P_{t-1}$, whereas K_t represents a simple technical index defined by (3), where P_{t-1}, P_{t-2} are day closing prices on day t-1 and t-2, respectively. We have implemented MS-Excel for calculation and the Table 1 gives a snippet of results.

$$\begin{aligned} K_t &= K_{t-1} Y_t, \quad \text{if } (P_{t-2} \le P_{t-1} \le P_t), \text{ or } (P_{t-2} \ge P_{t-1} \ge P_t) \text{ holds,} \\ K_t &= Y_t, \text{ otherwise.} \end{aligned} \tag{3}$$

t	1	2	3	4	5	6	7	8
P_t	904.0	866.7	893.7	903.6	866.5	825.5	775.3	810.2
Y_t		0.959	1.031	1.011	0.959	0.953	0.939	1.045
K_t		0.959	1.031	1.043	0.959	0.914	0.858	1.045

Table 1 Calculation of Y_t and K_t

Model 1: there is a basic model with just two states distinguishing either growth or decrease of Y_t. If Y_t <1 the state is denoted D, whereas for $Y_t \ge 1$ the corresponding state is denoted G. Processing Y_t yields the matrix $\boldsymbol{P}$ given by (4), which is not very interesting since both rows are just slightly different ones, hence for issuing trade signals not attractive as well.

$$\boldsymbol{P} = \begin{bmatrix} 0.481 & 0.519 \\ 0.443 & 0.557 \end{bmatrix} \tag{4}$$

Model 2: has eight states $\{D_4, D_3, D_2, D_1, G_1, G_2, G_3, G_4\}$ distinguishing different levels of growth and decrease of Y_t given by relations (5).

$$\begin{aligned} &D_4: Y_t < 0.97, \quad D_3: 0.97 \le Y_t < 0.98, \quad D_2: 0.98 \le Y_t < 0.99, \quad D_1: 0.99 \le Y_t < 1.00, \\ &G_1: 1.00 \le Y_t \le 1.01, \quad G_2: 1.01 < Y_t \le 1.02, \quad G_3: 1.02 < Y_t \le 1.03, \quad G_4: 1.03 < Y_t\,. \end{aligned} \tag{5}$$

Applying the same procedure coded in MS-Excel upon Y_t gives both the transition probability matrix $\boldsymbol{P}$ and conditional probabilities of decrease ${}_{\mathrm{D}}q_i = \sum_{j=1}^{4} p_{ij}$ and growth ${}_{\mathrm{G}}q_i = \sum_{j=5}^{8} p_{ij}$ depending upon states Di and Gi, i=1,.., 4, respectively, which constitutes the columns of matrix $\boldsymbol{Q}$. Both matrices are listed in (6). Searching for maximal values in columns of matrix $\boldsymbol{Q}$ we can conclude that the state D_3 provides 0.627 of probability to decrease, whereas the state G_4 provides 0.636 probability of growth.

$$\boldsymbol{P} = \begin{bmatrix} 0.180 & 0.040 & 0.120 & 0.120 & 0.140 & 0.100 & 0.120 & 0.180 \\ 0.102 & 0.102 & 0.102 & 0.322 & 0.202 & 0.085 & 0.068 & 0.017 \\ 0.052 & 0.059 & 0.105 & 0.241 & 0.261 & 0.183 & 0.092 & 0.007 \\ 0.026 & 0.049 & 0.112 & 0.282 & 0.375 & 0.114 & 0.028 & 0.014 \\ 0.015 & 0.019 & 0.094 & 0.322 & 0.388 & 0.114 & 0.029 & 0.019 \\ 0.016 & 0.021 & 0.091 & 0.289 & 0.375 & 0.144 & 0.027 & 0.037 \\ 0.031 & 0.062 & 0.092 & 0.322 & 0.277 & 0.092 & 0.062 & 0.062 \\ 0.058 & 0.058 & 0.104 & 0.104 & 0.267 & 0.149 & 0.134 & 0.126 \end{bmatrix}, \boldsymbol{Q} = \begin{bmatrix} 0.460 & 0.540 \\ 0.627 & 0.373 \\ 0.458 & 0.542 \\ 0.469 & 0.531 \\ 0.451 & 0.549 \\ 0.417 & 0.583 \\ 0.508 & 0.492 \\ 0.364 & 0.636 \end{bmatrix} \tag{6}$$

Model 3: has eight states denoted $\{D_4, D_3, D_2, D_1, G_1, G_2, G_3, G_4\}$ as well, however distinguishing different levels of growth and decrease of K_t, which are given by relations (7). Inspecting both the state space and K_t introduced we conclude that some of transition probabilities should be zero since such transitions are not possible.

$$\begin{aligned} &D_4: K_t < 0.97, \quad D_3: 0.97 \le K_t < 0.98, \quad D_2: 0.98 \le K_t < 0.99, \quad D_1: 0.99 \le K_t < 1.00, \\ &G_1: 1.00 \le K_t \le 1.01, \quad G_2: 1.01 < K_t \le 1.02, \quad G_3: 1.02 < K_t \le 1.03, \quad G_4: 1.03 < K_t . \end{aligned} \tag{7}$$

Now, analyzing K_t yields the following transition probability matrix $\boldsymbol{P}$ and matrix of conditional probabilities $\boldsymbol{Q}$, which are both given by (8).

$$\boldsymbol{P} = \begin{bmatrix} 0.510 & 0 & 0 & 0 & 0.176 & 0.157 & 0.085 & 0.072 \\ 0.462 & 0.103 & 0 & 0 & 0.269 & 0.115 & 0.038 & 0.013 \\ 0.098 & 0.172 & 0.202 & 0 & 0.307 & 0.129 & 0.074 & 0.018 \\ 0.024 & 0.071 & 0.178 & 0.171 & 0.411 & 0.111 & 0.027 & 0.007 \\ 0.011 & 0.030 & 0.100 & 0.300 & 0.185 & 0.300 & 0.041 & 0.033 \\ 0.009 & 0.018 & 0.080 & 0.298 & 0 & 0.248 & 0.218 & 0.129 \\ 0.023 & 0.023 & 0.094 & 0.297 & 0 & 0 & 0.250 & 0.313 \\ 0.042 & 0.031 & 0.104 & 0.318 & 0 & 0 & 0 & 0.505 \end{bmatrix}, \boldsymbol{Q} = \begin{bmatrix} 0.510 & 0.490 \\ 0.564 & 0.438 \\ 0.472 & 0.528 \\ 0.444 & 0.556 \\ 0.441 & 0.559 \\ 0.404 & 0.596 \\ 0.438 & 0.563 \\ 0.495 & 0.505 \end{bmatrix} \tag{8}$$

Since both conditional probabilities in matrix $\boldsymbol{Q}$ range between 40% and 60% there are not much interesting for trading signal issue.

Model 4: has similar state space as in the previous model. However, the algorithm is applied on filtered sequence of K_t states. Filtering, it means omitting subsequently repeated states. For example, let $D_2\, D_3\, D_3\, G_1\, G_1\, G_1\, G_2\, D_1\, D_1\, G_2$ be an un-filtered sequence, by filtering we get a filtered one $D_2\, D_3\, G_1\, G_2\, D_1\, G_2$. Hence filtering concerns trend changes within K_t states. Finally, analyzing filtered sequence of K_t states provides the transition probability matrix $\boldsymbol{P}$ and matrix of conditional probabilities $\boldsymbol{Q}$, which are both given by (9).

$$\boldsymbol{P} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0.360 & 0.320 & 0.173 & 0.147 \\ 0.514 & 0 & 0 & 0 & 0.300 & 0.129 & 0.043 & 0.014 \\ 0.123 & 0.215 & 0 & 0 & 0.385 & 0.162 & 0.092 & 0.023 \\ 0.028 & 0.085 & 0.215 & 0.000 & 0.497 & 0.134 & 0.033 & 0.008 \\ 0.014 & 0.036 & 0.123 & 0.368 & 0.000 & 0.368 & 0.050 & 0.041 \\ 0.012 & 0.024 & 0.107 & 0.395 & 0 & 0 & 0.290 & 0.172 \\ 0.031 & 0.031 & 0.125 & 0.396 & 0 & 0 & 0 & 0.417 \\ 0.084 & 0.063 & 0.211 & 0.642 & 0 & 0 & 0 & 0 \end{bmatrix}, \boldsymbol{Q} = \begin{bmatrix} 0 & 1 \\ 0.514 & 0.486 \\ 0.338 & 0.662 \\ 0.329 & 0.671 \\ 0.541 & 0.459 \\ 0.538 & 0.462 \\ 0.583 & 0.417 \\ 1 & 0 \end{bmatrix} \tag{9}$$

Searching for maximal values in the second column of matrix $\boldsymbol{Q}$ which gives us conditional probabilities for growth we find states D_1 and D_2 to be interesting since having 67.1% of growth in next day and 66.2%, respectively.

3 Numerical results

Having constructed transition probability matrices of our models we calculate development of vectors $\boldsymbol{p}(n)$ in time using recurrence equation (2) with starting state distribution given by $\boldsymbol{p}(1)$. Different vectors $\boldsymbol{p}(1)$ will produce different short-term evolutions of investigated systems modeled by our models based upon MC. We implement sw system Mathematica, Wolfram Research, Inc., to produce both numerical and graphical results.

First, we show transition probability matrices of the models 2-4 as bar-chart graphs. They are all produced by `BarChart3D` Mathematica commands and displayed on Figures 1, 2, and 3. As the models 3 and 4 are closely related as to their K_t state sequences, there is interesting to see differences in between. The model 3 handles with un-filtered sequence whereas the model 4 with the filtered one. The Figure 4 shows difference matrix $\boldsymbol{P}_3 - \boldsymbol{P}_4$ where $\boldsymbol{P}_3$ denotes matrix $\boldsymbol{P}$ of model 3 given by (8) and $\boldsymbol{P}_4$ stands for matrix $\boldsymbol{P}$ given by (9), respectively.

Calculation of vectors $\boldsymbol{p}(n)$ is realized by Mathematica notebook, too. The recurrence equation (2) is simply implemented by following commands: `pnkArr={q1};`

```
Do[(q2=q1.P;pnkArr=Append[pnkArr,q2];q1=q2,{k}];
```

where `pnkArr` is an array containing vectors $\boldsymbol{p}(n)$ for n starting from 1 till given k, e.g. k=8, lists `q1` and `q2` represent vectors $\boldsymbol{p}(n)$ and $\boldsymbol{p}(n+1)$ from (2), and variable `k` sets the number of repetitions within the loop `Do`.

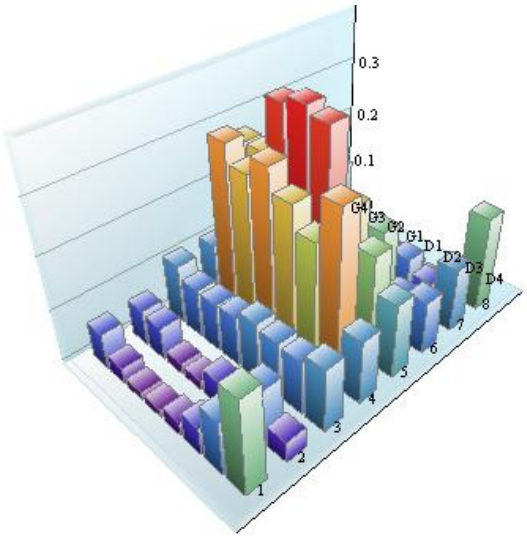


Figure 1 Matrix $\boldsymbol{P}$ of model 2

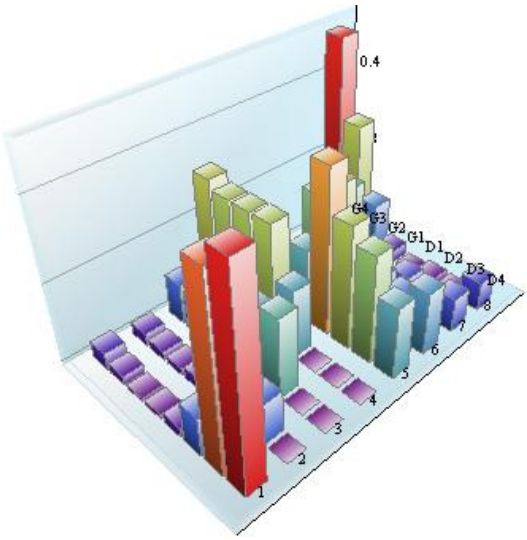


Figure 2 Matrix $\boldsymbol{P}$ of model 3

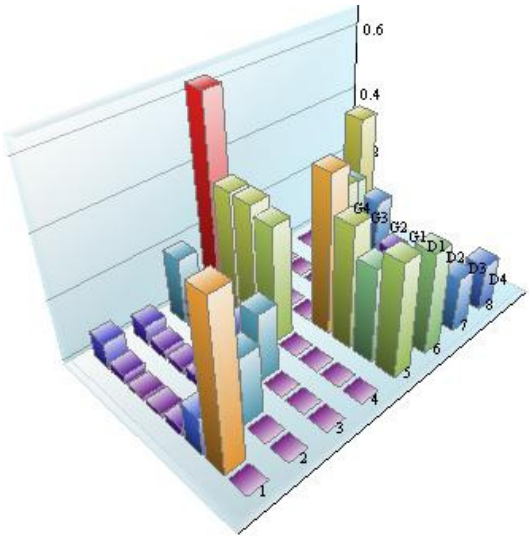


Figure 3 Matrix $\boldsymbol{P}$ of model 3

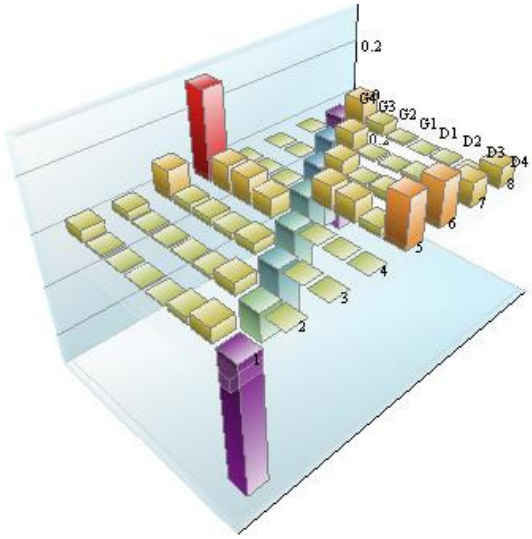


Figure 4 Difference matrix $\boldsymbol{P}_3 - \boldsymbol{P}_4$

In accordance with values of conditional probabilities of decrease ${}_{\mathrm{D}}q_i$ and growth ${}_{\mathrm{G}}q_i$ depending upon states Di and Gi, i=1,..,4, which constitutes the columns of matrices $\boldsymbol{Q}$ for each model being presented, we select initial vectors $\boldsymbol{p}(1)$. The calculated probabilities of states are denoted systematically $p(\mathrm{D}_4)$, $p(\mathrm{D}_3)$, .., $p(\mathrm{G}_3)$, $p(\mathrm{G}_4)$ are summarized in following tables and discussed further.

The model 2 has provided state G_4 with the greatest probability of growth and state D_3 with the greatest probability of decrease. Some results are summarized in Table 2, and corresponding graphs for n = 1,..,8 are given on Figure 5 and 6. When starting with $p(\mathrm{G}_4)$=1, the maximal conditional probabilities are located at the state G_1. The similar results hold for initial vector $\boldsymbol{p}(1)$ having $p(\mathrm{D}_3)$=1, too.

n	$p(\mathrm{D}_4)$	$p(\mathrm{D}_3)$	$p(\mathrm{D}_2)$	$p(\mathrm{D}_1)$	$p(\mathrm{G}_1)$	$p(\mathrm{G}_2)$	$p(\mathrm{G}_3)$	$p(\mathrm{G}_4)$
1	0	0	0	0	0	0	0	1
2	0.058	0.058	0.104	0.104	0.267	0.149	0.134	0.126
3	0.0423	0.0433	0.0995	0.2653	0.3162	0.1246	0.0604	0.0484
4	0.0355	0.0402	0.1013	0.2801	0.3392	0.1231	0.0474	0.0332

n	$p(D_4)$	$p(D_3)$	$p(D_2)$	$p(D_1)$	$p(G_1)$	$p(G_2)$	$p(G_3)$	$p(G_4)$
1	0	1	0	0	0	0	0	0
2	0.102	0.102	0.102	0.322	0.202	0.085	0.068	0.017
3	0.0500	0.0472	0.1043	0.2739	0.3160	0.1184	0.0524	0.0387
4	0.0361	0.0407	0.1012	0.2809	0.3388	0.1234	0.0466	0.0323

Table 2 Model 2 ~ vectors $\boldsymbol{p}(n)$, n=1,..,4 for two different vectors $\boldsymbol{p}(1)$

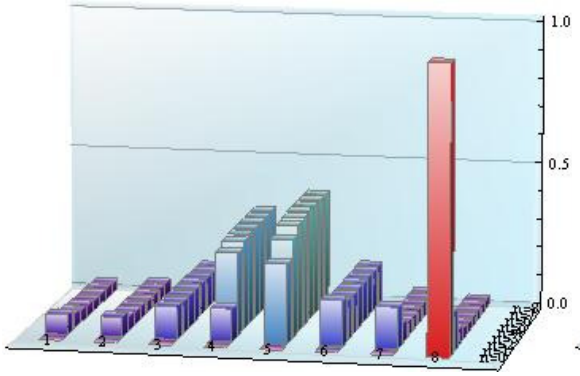


Figure 5 Model 2 ~ $\boldsymbol{p}(n)$ for $\boldsymbol{p}(1)$ with $p(G_4)$=1

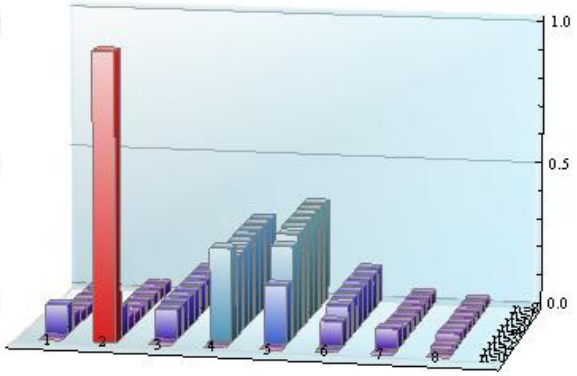


Figure 6 Model 2 ~ $\boldsymbol{p}(n)$ for $\boldsymbol{p}(1)$ with $p(D_3)$=1

We have already noted that the model 4 provide states D_1 and D_2 to be interesting ones for growth in the next day. We make similar short-term analysis as with the previous model 2, and some results are summarized in Table 3. The corresponding graphs for n = 1,..,8 are depicted on Figure 7 and 8. When starting with $p(D_1)$=1, the maximal conditional probability migrates for steps n=2, 3 and 4 to states G_1, D_1 and D_1, respectively. The similar results hold for initial vector $\boldsymbol{p}(1)$ having $p(D_2)$=1, too. So, it reflects the fact that the most frequent change from decrease to growth in the filtered sequence of K_t states is the transition $D_1 \rightarrow G_1$ with its conditional probability 0.497.

n	$p(D_4)$	$p(D_3)$	$p(D_2)$	$p(D_1)$	$p(G_1)$	$p(G_2)$	$p(G_3)$	$p(G_4)$
1	0	0	0	1	0	0	0	0
2	0.028	0.085	0.215	0.000	0.497	0.134	0.033	0.008
3	0.0804	0.0689	0.0813	0.2540	0.1184	0.2377	0.0919	0.0674
4	0.0655	0.0561	0.1203	0.2172	0.2071	0.1254	0.1076	0.1008

n	$p(D_4)$	$p(D_3)$	$p(D_2)$	$p(D_1)$	$p(G_1)$	$p(G_2)$	$p(G_3)$	$p(G_4)$
1	0	0	1	0	0	0	0	0
2	0.123	0.215	0.000	0.000	0.385	0.162	0.092	0.023
3	0.1226	0.0220	0.0810	0.2569	0.1088	0.2088	0.0968	0.1031
4	0.0442	0.0576	0.1248	0.2270	0.2096	0.1297	0.1041	0.1030

Table 3 Model 4 ~ vectors $\boldsymbol{p}(n)$, n=1,..,4 for two different vectors $\boldsymbol{p}(1)$

The Mathematica command used for pictures export is `Export[file_name.ext,img]`, where `ext` defines a graphical format desired, e.g. jpeg, and `img` holds the picture created, e.g. `lp07m4pn`. This picture is created by versatile command `BarChart3D` again, e.g. `lp07m4pn=BarChart3D[pnkArr,options]`, where `options` may express bar spacing, labels, etc.

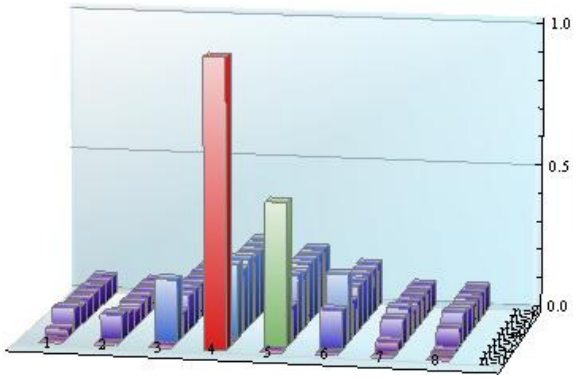


Figure 7 Model 4 ~ $\boldsymbol{p}(n)$ for $\boldsymbol{p}(1)$ with $p(D_1)=1$

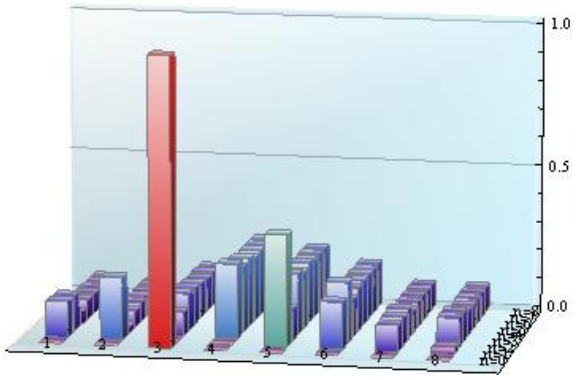


Figure 8 Model 4 ~ $\boldsymbol{p}(n)$ for $\boldsymbol{p}(1)$ with $p(D_2)=1$

Sure, there is possible to analyze other investing strategies just by setting different initial vectors $\boldsymbol{p}(1)$, as well. However, such analysis is out of our present paper.

4 Conclusion

We have proposed and discussed some results of four models based upon MCA with time independent transition probability matrices to analyze and predict trends of Prague stock exchange index PX. In further research we would like to concentrate ourselves to three main challenging directions

- construction of different state space S within MCA,
- problem-oriented filtering procedures of trade sequences,
- development and empirical evaluation of various technical indices for TA and implementation of non-homogeneous MC.

Acknowledgements

Authors acknowledge gratefully funding by the internal project SGS_2012–036 provided at the University of West Bohemia.

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), 1731–1764.
[2] Doubleday, K., J. and Esunge, J., N.: Application of Markov Chains to Stock Trends. *Journal of Mathematics and Statistics* **7** (2), 2011, 103-106.
[3] Fama, E., F.: Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance* **25** (1970), 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] Sweeny, R., J.: Some new filter rule tests: methods and results and Quantitative Analysis. *Journal of Financial Economics* **23** (1988), 285–300.
[7] Vasanthi, S., Subha M., V. and Nambi S., T.: An empirical Study on Stock Index trend prediction using Markov Chain analysis. *JBFSIR* **1**(2011), 72 – 90.

Voracity effect and Wagner's law in the PIIGS

Irena Szarowská[1]

Abstract. The paper provides direct empirical evidence on cyclicality and the short-term and long-term relationship between government expenditure and output in the PIIGS (Portugal, Ireland, Italy, Greece, Spain) in a period 1990-2010. We follow Akitoby et al. and apply Johansen cointegration test and the error correction model on annul data of GDP and government expenditure in compliance with the COFOG international standard. The government expenditure functions are procyclical in most PIIGS countries (68% cases in the sample). Output and government expenditure are cointegrated for at least six of the expenditure categories in every country and it implies a long-term relationship between government expenditure and output consistent with Wagner's law. Average value of long-run elasticity coefficients is 1.30 for all expenditure functions, 1.17 for total government expenditure. The values of the coefficients for the short-run relationship between expenditure and output confirm the voracity hypothesis, as they suggest that in response to a given shock to real GDP, government expenditure rises by even more in percentage points.

Keywords: government expenditure, cyclicality, voracity effect, long-run elasticity, short-run elasticity.

JEL Classification: C32, H50, E62
AMS Classification: 90C15

1 Introduction

Government expenditure and factors of their growth are a serious problem of many countries. As Mutascu and Milos [16] mention, the economic theory provides two main categories of arguments that explain the public sector size in time and among countries. The first category has as starting point the Wagner law, according to which the elasticity of government expenditure compared to GDP is greater than one. As countries become more developed, the demand for public goods raises and is consistent with the increasing ability to collect the necessary funds. On the other hand, the "Baumol cost disease", explains that the percentage of government expenditure increases because the raise of public servants' salaries is higher than their productivity, while the price related to public services demand is relatively non)elastic. The second category of arguments is political. For election purposes, the fiscal policy, especially those concerning the government expenditure tends to be inconsistent in time and focuses on greater deficits and greater public sectors.

We can find a view that government expenditure should act as a stabilizing force and move in a countercyclical direction. Contrary to the theory (it implies that government expenditure is countercyclical), many of empirical studies found evidence that government expenditure is procyclical. See Hercowitz and Strawczynski [11], Kaminsky et al. [13], Alesina et al. [3], Rajkumar and Swaroop [18], Hamerníková [10], Ganelli [7] or Szarowská [19] for more details. Talvi and Vegh [20] show that fiscal procyclicality is evident in a much wider sample of countries. Analysis of Lane [14] finds procyclicality in a single-country time series study of Irish fiscal policy. Lane [15] also shows that the level of cyclicality varies across expenditure categories and across OECD countries. Abbott and Jones [1] test differences in the cyclicality of government expenditure across functional categories. Their evidence from 20 OECD countries suggests that procyclicality is more likely in smaller functional budgets, but capital expenditure is more likely to be procyclical for the larger expenditure categories. Many of researches as Gavin et al. [8], Gavin and Perotti [9] focused on Latin America. On the one hand, Galí [6] shows in his research that expenditure is countercyclical. However, other papers show no discernible pattern. Fiorito and Kollintzas [5] document for G7 countries, the correlation between government consumption and output indeed appears to show no pattern and be clustered around zero. The differences in these results depend on the components of expenditure being measured. Government transfers and subsidies are found to have become substantially more countercyclical.

Economic performance is greatly influenced by the level and the structure of government expenditure. It is not only a potential automatic stabilizer, but it is also a tool of political actions. In fact, development of government expenditure is often associated with Wagner ´s law and voracity effect. Wagner's law states that government activity increases as economies grow, with the pace of increase being different for different branches of

[1] Silesian University/School of Business Administration, Department of Finance, Univerzitni nam. 1934, 73340 Karvina, Czech Republic, szarowska@opf.slu.cz.

government. Voracity effect occurs if a positive shock to income leads to a more than proportional increase in public expenditure, even if the shock is expected to be temporary. The voracity is usually attributed to weak institutions and ethnic fractionalization, manifested in the presence of multiple interest groups seeking to secure a greater share of national wealth by demanding larger public expenditure on their behalf. The existing literature testing Wagner's law varies considerably in terms of the dependent and independent variables chosen to "test" the law. Wagner originally proposed that as industrialization or social progress proceeded, public sectors would grow in relative importance. In practice, researchers use different measures of national income as a measure of this social progress. Peacock and Scott [17] point out on the fact that there are at least 14 different measures of government expenditure that have been used in the literature, and at least 13 different measures of output, including output per capita. In this paper we adopt the simplest formulation of Wagner's law by focusing on the relationship between aggregate economic activity and government expenditure in compliance with the COFOG international standard. Most studies analyzing the cyclicality of government expenditure and output have used a panel data methodology that has not fully exploited the time-series properties of the data. On the other hand, studies testing for a long-run relationship, such as Wagner's law, have ignored the short-term aspects of this relationship. In the literature on cyclicality, many studies use panel data models that are not well suited to exploring short-term versus long-term relationships. We exploit both the time-series and cross-sectional aspects using an error-correction framework.

The aim of the paper is to provide direct empirical evidence on cyclicality and the short-term and the long-term relationship between government expenditure and output in five selected European countries. Although the theory implies that government expenditure is countercyclical, recent evidence suggests that it is procyclical. Previously published studies are weakly supported by the data from PIIGS countries in which results can vary. We follow Akitoby et al. [2] and apply Johansen cointegration test [12] and the error correction model on annul data of GDP and government expenditure during 1990–2010. The article is organized as follows. In the next section, we describe the dataset and empirical techniques used. Then, we present the results of government expenditure cyclicality and long-run and short-run relationship between output and government expenditure. We conclude with a summary of key findings.

2 Data and methodology

The dataset consists of annual data on GDP and government expenditure in compliance with the COFOG international standard during the period 1990–2010. It is not possible to use higher frequently time series data as COFOG classification analyzes and reports only annual data. We use data about expenditure and output from Eurostat and OECD database, GDP deflators in 2005 values are taken from the World Bank. The countries included in the analysis are Portugal, Ireland, Italy, Greece and Spain. The series for GDP and total government expenditure and its subcomponent are adjusted at constant prices. In line with Akitoby et al. [2], we investigated fiscal and output co-movements by the approach proposed by Lane [15]. We estimated the elasticity of government expenditure with respect to output, based on country-by-country time-series regressions. Next we used an error-correction approach, which allows us to distinguish between the short-term effect of output on government expenditure and any longer-term effect between these two variables. Most of the results were calculated in econometric program Eviews 7.

Many studies point out that using non-stationary macroeconomic variable in time series analysis causes superiority problems in regression. Thus, a unit root test should precede any empirical study employing such variables. We decided to make the decision on the existence of a unit root through Augmented Dickey–Fuller test (ADF test). The equation (1) is formulated for the stationary testing.

$$\Delta x_t = \delta_0 + \delta_1 t + \delta_2 x_{t-1} + \sum_{i=1}^{k} \alpha_i \Delta x_{t-i} + u_t \tag{1}$$

ADF test is used to determine a unit root x_t at all variables in the time t. Variable Δx_{t-i} expresses the lagged first difference and u_t estimate autocorrelation error. Coefficients δ_0, δ_1, δ_2 and α_i are estimated. Zero and the alternative hypothesis for the existence of a unit root in the x_t variable are specified in (2).

$$H_0: \delta_2 = 0,\ H_\varepsilon: \delta_2 < 0 \tag{2}$$

The result of ADF test confirm the stationary of all time series at the first difference. Testing the stationary is the essential assumption for implementation of cointegration approach. It is necessary to confirm that time series are non-stationary at level data but stationary at first difference.

We suppose there is a steady-state relationship between government expenditure and output given by (3).

$$G = AY^{\delta} \tag{3}$$

G represents government expenditure, Y means output and Eq. (3) can also be written in linear form:

$$logG = a + \delta logY,\ a = logA \tag{4}$$

If the adjustment of expenditure G to its steady-state $\bar{G}$ is gradual, then the level of expenditure will respond to transitory changes in output, and G will move gradually toward its steady-state, or equilibrium level. To capture this gradual move, we specify a general autoregressive distributed lag specification for expenditure category i in period t:

$$logG_{it} = \mu + \alpha logG_{it-1} + \beta_0 logY_t + \beta_1 logY_{t-1} + \varepsilon_t, \left|\alpha\right| < 1 \quad (5)$$

We can solve for the static, steady-state equilibrium by assuming that output is at its steady-state level $\bar{Y}$ and ignoring the error term:

$$log\bar{G} = \frac{\mu}{1-\alpha} + \frac{\beta_0+\beta_1}{1-\alpha} log\bar{Y},\ \delta = 1-\alpha \quad (6)$$

More generally, we could allow output to grow at rate g. In this case, the only difference is that the constant term becomes $\frac{\mu+(\beta_0-\delta)g}{1-\alpha}$, which depends on g. To reflect the steady state, (5) can be rearranged as the error correction model (7).

$$logG_{it} = \mu + \beta_0 logY_t + \gamma(logG_{it-1} - \delta logY_{t-1}) + \varepsilon_t \quad (7)$$

In (7), we can interpret $\beta_0 \Delta logY_t$ as the short-term impact of output on expenditure and β_0 as the short-run elasticity of government expenditure with respect to output. The error correction term $\gamma(logG_{it-1} - \delta logY_{t-1})$ captures deviations from the steady-state, or long-run equilibrium, where δ is the long-run elasticity of government expenditure with respect to output, and γ is the rate at which government expenditure adjusts to past disequilibrium. μ is constants of the model, ε_t means residual component of long-term relationship.

Moreover, (7) can be rewritten as (8) and then used to test if there is a long-run relationship between government expenditure and output. In particular, following [4], if γ is significantly different from zero in (8), then output and government expenditure are cointegrated.

$$logG_{it} = \mu + \beta_0 logY_t + \gamma logG_{it-1} - \varphi logY_{t-1} + \varepsilon_t \quad (8)$$

where $\varphi = \gamma\delta$. The above derivation makes clear the underlying assumption that there is a elasticity relationship between output and expenditure, while the transitory deviations are random.

3 Structure and cyclicality of government expenditure

Government expenditure can help in overcoming the inefficiencies of the market system in the allocation of economic resources. It also can help in smoothing out cyclical fluctuations in the economy and influences a level of employment and price stability. We used government expenditure in compliance with the COFOG international standard (Classification of the Functions of Government) in our analysis. Total government expenditure is divided into 10 basic divisions:

- CF01: General public services
- CF02: Defense
- CF03: Public order and safety
- CF04: Economic affairs
- CF05: Environment protection
- CF06: Housing and community amenities
- CF07: Health
- CF08: Recreation; culture and religion
- CF09: Education
- CF10: Social protection

3.1 The structure of government expenditure

Firstly we analyzed the structure of government expenditure in a period 1995–2010. Results in Table 1 show the average share of government expenditure by functions on total expenditure and the average share of total government expenditure on GDP in each country during the analyzed period.

	CF01	CF02	CF03	CF04	CF05	CF06	CF07	CF08	CF09	CF10	Gtotal /GDP
Greece	23.86%	6.11%	2.58%	11.19%	1.17%	0.81%	11.50%	0.85%	7.19%	34.73%	46.43%
Spain	13.95%	2.77%	4.67%	12.03%	2.13%	2.53%	13.63%	3.51%	10.99%	33.80%	40.85%
Ireland	10.72%	1.61%	4.51%	13.70%	2.34%	4.33%	17.64%	1.81%	13.17%	30.18%	38.05%
Italy	21.66%	2.67%	3.99%	8.58%	1.65%	1.72%	12.73%	1.71%	9.47%	35.81%	50.13%
Portugal	15.18%	3.29%	4.21%	10.42%	1.43%	1.78%	14.72%	2.68%	14.50%	31.81%	43.77%
Average	17.08%	3.29%	3.99%	11.18%	1.74%	2.24%	14.04%	2.11%	11.06%	33.27%	43.85%

Table 1 Government expenditure - COFOG classification (in % of total G)

Five expenditure functions (Social protection, Economic affairs, Health, General public services and Education), on average, account nearly 87% of the total expenditure. The Social protection expenditure (CF10) is the

highest expenditure function in every country and it takes the third of total government expenditure. It contains, for example, expenditure on sickness and disability, old age, survivors, family and children, unemployment, housing, social exclusion and R&D social protection. The value of General public services (CF01) is the second highest category (17%). We can find the highest value of CF01 (23.86%) in Greece, followed by Italy (21.66%); it is due to a high expenditure on public debt services. On the other hand, Ireland has the smallest CF01 expenditure (less than 11%). Economics affairs (CF04) and (CF09) Education expenditure are in average very similar (11.18% resp. 11.06%), but the share differs in each country. Education expenditure are twice as high in Portugal as in Greece. The value of total government expenditure is the smallest in Ireland (38.05% GDP), the highest in Italy (50.13% GDP), and the average of all countries is 43.77% GDP; that expresses significant differences in size and importance of public sector in the sample of countries.

3.2 The cyclicality of government expenditure

As was already noted, government expenditure is a possible automatic stabilizer. The cyclicality of government expenditure is typically defined in terms of how expenditure moves with the output gap. If government expenditure increases when there is a positive output gap (i.e. output is below its potential), then expenditure is counter-cyclical. If potential output were observable or easy to estimate, one could define counter-cyclicality as above-average expenditure to output ratio whenever output was below its potential. As Akitoby et al. [2] mention, measuring potential output is difficult. As a consequence, it is not easy to discuss business cycles or cyclicality per se. Therefore we focus on co-movements of government expenditure and output as a proxy for cyclicality.

	G total	CF01	CF02	CF03	CF04	CF05	CF06	CF07	CF08	CF09	CF10
Greece	-0.91*	0.32*	-0.60*	-0.25	-1.28*	-0.01	-0.49	0.03	-0.80*	-1.10*	-0.39*
	(0.45)	(0.09)	(0.22)	(0.34)	(0.30)	(0.12)	(0.31)	(0.11)	(0.28)	(0.45)	(0.20)
Spain	0.23*	0.37*	-0.37*	0.41	-1.18*	-0.29	-0.17	-0.17*	-0.35	-0.04*	0.05*
	(0.06)	(0.15)	(0.19)	(0.41)	(0.34	(0.21)	(0.32)	(0.07)	(0.25)	(0.02)	(0.03
Ireland	0.29	-0.01*	-0.32*	0.23*	1.41	-0.04	-0.91*	-0.18**	0.28	-0.06*	-0.28*
	(0.18)	(0.00)	(0.14)	(0.12)	(1.03)	(0.05)	(0.39)	(0.11)	(0.36)	(0.02)	(0.07)
Italy	-0.00	-0.56**	-0.27*	-0.07*	-0.62*	-0.39*	-0.66*	-0.31**	-0.13	-0.31*	-0.34**
	(0.03)	(0.18)	(0.11)	(0.34)	(0.24)	(0.16)	(0.30)	(0.08)	(0.24)	(0.11)	(0.09)
Portugal	-0.01	-0.69**	-1.11*	-0.06	-0.40	-0.23*	-0.52*	-0.02	-0.48*	-0.56*	-0.16**
	(0.11)	(0.14)	(0.37)	(0.19)	(0.43)	(0.13)	(0.14)	(0.23)	(0.24)	(0.23)	(0.05)
Average	0.57	0.39	0.53	0.23	1.2	0.31	0.69	0.22	0.64	0.41	0.24
Share significant	40%	100%	100%	20%	60%	40%	60%	60%	40%	100%	100%

Table 2 The value of adjustment coefficient γ

Symbols *and ** denote significance at the 5% and 1% level, standard deviation are in parenthesis. Average means the average absolute values of significant coefficients only. Share significant means share of significant cases. Table 2 reports the estimates of the adjustment coefficient γ from equation (7), which is estimated by OLS with a correction for an autoregressive error term. γ is the rate at which government expenditure adjusts to past disequilibrium. In cases where γ is significant, we can conclude there is a cointegrating relationship between government expenditure and output. The results indicate significant difference across expenditure functions. There is a long-term relationship between total government expenditure and output consistent with Wagner's law, the share of significant results is 68% for all categories in all countries. Although the error correction term is not significant for all expenditure functions in any country of the sample, all countries have a significant error correction term for at least six of the expenditure functions (six in Greece and Spain, seven in Ireland and Portugal and eight in Italy). Moreover, the error correction term for General public services (CF01), Defense (CF02) Education (CF09) and Social protection (CF10) are significant in all countries. As expected, the adjustment coefficients are mostly negative (in 86% of cases), indicating dynamic stability. The implication of a significant error correction term is that there is in fact a long-term relationship between government expenditure and output. But it is suitable to point out that the existence of cointegration does not imply causality, which is consistent with Wagner's view that there is not necessarily a cause and effect relationship between economic development and government activity.

Table 3 summarizes the results about the long- run elasticity of expenditure with respect to output. Symbols *and ** denote significance at the 5% and 1% level, standard deviation are in parenthesis. Average means the average absolute values of significant coefficients only. Share significant means share of significant cases. It contains only significant coefficients; the long-run elasticity coefficient δ is significant in 84% cases. A positive value of δ is consistent with a wider interpretation of Wagner's law, as it implies that government expenditure

rises with national income. If δ is higher than one then this would be consistent with a narrow interpretation of Wagner's law, where government expenditure rises faster than national income.

	G total	CF01	CF02	CF03	CF04	CF05	CF06	CF07	CF08	CF09	CF10
Greece	1.02*	-1.50*	1.43*	4.20**	0.88**	0.58	1.19**	0.76**	2.69*	2.07**	0.78**
	(0.06)	(0.27)	(0.55)	(0.32)	(0.14)	(0.01)	(0.23)	(0.01)	(0.30)	(0.15)	(0.09)
Spain	-0.06	-0.65**	0.74**	0.81**	1.23**	1.37**	-0.20	1.94**	0.98**	2.85*	0.91**
	(0.13)	(0.14)	(0.09)	(0.07)	(0.09)	(0.07)	(0.32)	(0.33)	(0.10)	(0.93)	(0.01)
Ireland	0.36*	1.38**	0.55**	0.14	0.56**	0.13	1.11**	1.20**	0.79**	2.46**	0.86**
	(0.15)	(0.21)	(0.01)	(0.14)	(0.08)	(0.59)	(0.09)	(0.05)	(0.09)	(0.61)	(0.01)
Italy	0.94**	-1.97**	1.36*	3.13**	0.77**	1.66**	0.58**	2.37**	1.96**	0.78**	0.88**
	(0.01)	(0.18)	(0.62)	(0.41)	(0.01)	(0.18)	(0.00)	(0.27)	(0.21)	(0.00)	(0.14)
Portugal	2.34**	0.77**	0.64*	0.68**	0.73**	-0.22**	0.58**	2.55**	0.63**	0.77**	0.89**
	(0.40)	(0.01)	(0.0)	(0.02)	(0.00)	(0.72	(0.01	(0.25)	(0.00)	(0.00)	(0.02)
Average	1.17	1.26	0.95	2.20	0.83	1.08	0.86	1.76	1.41	1.79	0.86
Share significant	80%	100%	100%	80%	100%	60%	80%	100%	100%	100%	100%

Table 3 The long-run elasticity coefficient δ

The long-term elasticity of government expenditure and output δ is mostly positive (in 92% of cases), and it is the highest for Public order and safety (CF03) due to the extremely high δ in Italy (it greatly increased the average). Moreover, δ is for total expenditure larger than one (1.17), average value is 1.30 for all expenditure functions. It is consistent with the narrow interpretation of Wagner's law and indicating that in the long-term, the public sector is increasing in relative importance. The coefficient for long-run elasticity was significant in all countries for all expenditure functions with the exception of Public order and safety (CF03), Environment protection (CF05) and Housing and community amenities (CF06). In Table 3, we can also find the long-run δ lower than one, it means that the expenditure function rises slower than national income.

	G total	CF01	CF02	CF03	CF04	CF05	CF06	CF07	CF08	CF09	CF10
Greece	-0.54	2.34*	5.97*	4.23*	0.96	1.92*	0.96	3.45*	-4.02	0.12	0.47
	(1.14)	(0.83)	(2.07)	(2.36)	(1.73)	(0.80)	(1.38)	(1.33)	(2.56)	(1.62)	(0.62)
Spain	1.21**	1.01*	0.19	2.11*	-0.29	-0.65	0.38	0.79*	-0.29	0.89**	1.21*
	(0.21)	(0.54)	(0.34)	(1.10)	(0.76)	(0.88)	(2.28)	(0.33)	(1.24)	(0.22)	(0.51
Ireland	-0.20	-0.63	0.83*	1.39*	1.11	1.43*	-1.65	-1.25*	2.92*	0.55*	-1.44*
	(0.70)	(0.39)	(0.31)	(0.48)	(4.68)	(0.60)	(1.52)	(0.49)	(1.50)	(0.15)	(0.58)
Italy	0.44*	1.05*	-0.43	0.18	0.52	0.64	-0.35	-0.55	1.14*	0.60*	-0.67**
	(0.23)	(0.50)	(0.77)	(0.89)	(1.53	(0.38)	(5.01)	(0.36)	(0.55)	(0.27)	(0.22)
Portugal	0.07	-0.69**	1.00	-2.63*	0.49	0.19	4.38*	1.13*	0.49	0.42	-1.34*
	(0.35)	(0.14)	(0.76)	(1.24)	(1.14)	(0.91)	(1.30)	(0.57)	(0.84)	(0.88)	(0.69)
Average	0.83	1.27	3.40	2.59	-	1.68	4.38	1.65	2.3	0.68	1.16
Share significant	40%	80%	40%	80%	0%	40%	20%	80%	40%	60%	80%

Table 4 The short-run elasticity coefficient β

Table 4 summarizes results about the short-run elasticity of expenditure with respect to output. Symbols *and ** again denote significance at the 5% and 1% level, standard deviation are in parenthesis. Average means the average absolute values of significant coefficients only. The results and conclusions for the short-run elasticity are not so unequivocal. For all expenditure categories, the average coefficient is 2.09. Although the short-run elasticity is positive for 79% of the cases in the sample, it´s needed to points out on 51% statistical significant of results. However, the coefficient value above one is consistent with the voracity hypothesis, as it suggests that in response to a given shock to real GDP, government expenditure rises by even more in percentage points.

4 Conclusion

The aim of this article was to provide direct empirical evidence on cyclicality and the long-term and short-term relationship between government expenditure and output in five selected European countries (namely Portugal, Ireland, Italy, Greece and Spain) in a period 1990–2010. We analyzed annual data on government expenditure in compliance with the COFOG international standard. Although the theory implies that government ex-

penditure is countercyclical, our research does not prove that. The results confirm cyclical development of government expenditure on GDP, Wagner´s law and voracity effect in PIIGS countries during 1990–2010.

We used Johansen cointegration test and the error correction model. Output and government expenditure are cointegrated for at least six of the expenditure functions in every country and it implies a long-term relationship between government expenditure and output. The government expenditure functions are procyclical in most countries (68% cases in the sample). Average value of long-run elasticity coefficient is 1.30 for all expenditure functions, 1.17 for total government expenditure. It is consistent with the interpretation of Wagner's law and indicates that the public sector is increasing in relative importance in the long-term. The δ coefficient was significant in all countries for all expenditure functions with the exception of CF03, CF05 and CF06.

We also analyzed the short-run relationship between expenditure and output. Results are not unambiguous due to a relatively low statistical significance (51%). However, the coefficient values (average is 2.09) confirm the voracity hypothesis, as they suggest that in response to a given shock to real GDP, government expenditure rises by even more in percentage points.

Acknowledgements

This paper ensued thanks to the support of the grant GAČR 403/11/2073 "Procyclicality of financial markets, asset price bubbles and macroprudential regulation".

References

[1] Abbott, A., and Jones, P.: Procyclical government spending: Patterns of pressure and prudence in the OECD. *Economics Letters* **111** (2011), 230–232.
[2] Akitoby, B., Clements, B., Gupta, S., and Inchauste, G.: Public spending, voracity, and Wagner's law in developing countries. *European Journal of Political Economy* **22** (2008), 908–924.
[3] Alesina, A., Campante, F. R., and Tabellini, G.: Why is fiscal policy often procyclical? *Journal of the European Economic Association* **6** (2008), 1006–1030.
[4] Ericsson, N., and Mckinnon, J.: Finite sample properties of error correction tests for cointegration. *Econometrics Journal* **5** (2002), 285–318.
[5] Fiorito, R., and Kollintzas, T.: Stylized facts of business cycles in the G7 from a real business cycles perspective. *European Economic Review* **38** (1994), 235–269.
[6] Galí, J.: Government size and macroeconomic stability. *European Economic Review* **38** (1994), 117–132.
[7] Ganelli, G.: The International Effects of Government Spending Composition. *Economic Modelling* **27** (2010), 631–640.
[8] Gavin, M., Hausmann, R., Perotti, R., and Talvi E.: *Managing Fiscal Policy in Latin America and the Caribbean: Volatility, Procyclicality, and Limited Creditworthiness*. IDB Working Paper No. 326, 1996.
[9] Gavin, M., and Perotti, R.: Fiscal policy in Latin America. *Macroeconomics Annual* **12** (1997), 11–70.
[10] Hamerníková, B. : Komparace vývoje veřejných výdajů ve vybraných zemích EU [online]. 2009 [cit 2010-11-10]. Url: http://kvf.vse.cz/wp-content/uploads/2010/06/1239726368_sb_hamernikovamaaytovaullikova.pdf.
[11] Hercowitz, Z. and Strawczynski, M.: Cyclical Ratcheting in Government Spending: Evidence from the OECD. *Review of Economics and Statistics* **86** (2004), 353–361.
[12] Johansen, S.: Cointegration and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica* **59** (1991), 1551–1580.
[13] Kaminsky, G. L., Reinhart, C. M., and Vegh, C. A.: When it Rains, it Pours: Procyclical Capital Flows and Macroeconomic Policies. *Nber Macoeconomics Annual* **1** (2004), 11–53.
[14] Lane, P. R.: International Perspectives on the Irish Economy. *Economic and Social Review* **29** (1998), 217–222.
[15] Lane, P. R.: The Cyclical Behaviour of Fiscal Policy: Evidence from the OECD. *Journal of Public Economics* **87** (2003), 2661–2675.
[16] Mutascu, M., and Milos, M.: Optimal Size of Government Spending. The Case of European Union Member States. *Annales Universitatis Apulensis Series Oeconomica* **11** (2009), 447–456.
[17] Peacock, A., and Scott, A.: The curious attraction of Wagner's law. *Public Choice* **102** (2000), 1–17.
[18] Rajkumar, A. S., and Swaroop, V.: Public Spending and Outcomes: Does Governance Matter? *Journal of Development Economics* **86** (2008), 96–111.
[19] Szarowská, I.: Public Spending Cyclicality and Wagner's Law in Central and Eastern European Countries. *Acta Universitatis Agriculturae et Silviculturae Mendeleianae Brunensis* **60** (2012), 383–390.
[20] Talvi, E., and Vegh, C. A.: Tax base variability and procyclical fiscal policy in developing countries. *Journal of Development Economics* **78** (2005), 156–190.

Estimation of value of travel time savings using conditional Logit model

Petr Šenk[1], Stanislav Biler[2]

Abstract. This paper presents results of study focused on estimation of value of travel time savings in the context of leisure travel. The study refers to the concept of willingness-to-pay for reduction of travel time. Data on choice behavior were collected by means of stated preference survey on representative sample of population in the Czech Republic and analyzed using methods of discrete choice analysis.
Results of the study suggest that, in case of the most common type of leisure trips – weekend trips, estimated values of travel time savings obtained by the above mentioned methods are consistent with values recommended by literature, i.e. within the range of 25% and 50% of the average hourly wage.

Keywords: value of travel time savings, conditional logit model, stated preference survey, willingness-to-pay.

JEL Classification: R41, C25
AMS Classification: 91B06

1 Introduction

Value of travel time savings (VTTS) is one of the key inputs to transport planning models and tools for management and appraisal of road investments decisions. Information on VTTS is essential for parameterization of destination, mode, route and departure time choice components of complex travel demand models. This paper describes methodology and results of study aimed at valuation of travel time in the Czech Republic in the context of leisure travel by car.

Literature on estimation of VTTS refers to two main approaches, whose adequacy differs according to the purpose of travel [1]. So called cost saving approach is being used in case of business trips, during which the travel time of employees is viewed as unproductive. For such type of trips the VTTS is equal to the monetary valuation of employees' productive output. Cost saving approach is recommended also in case of commercial goods traffic, where the main trip characteristics (for example destination, route, and departure time) are given by the business policy of employer. On the contrary, non-work trips, including commuting, shopping or leisure trips, are to a large extent determined by the driver himself. Is such case it is appropriate to focus directly on preferences of drivers and their willingness-to-pay for reduction of travel time.

Section 2 presents in details the method of willingness-to-pay that combines stated preference method for data collection and discrete choice modeling for data analysis. Section 3 provides description of the data collection method and key descriptive statistics of the working dataset. Section 4 presents resulting values of travel time savings and section 5 discussion of results.

2 Methodology

2.1 Discrete choice model

In the study, valuation of travel time savings refers to basic economical concepts of utility theory and theory of rational choice. According to these theories consumer (in our case driver as user of transport infrastructure) choses from the finite set of all available options such that brings him or her the highest utility. Utility is understood as linear combination of attributes assigned to each option (travel time, comfort, fuel consumption, etc.) and individual preference weights that individuals assign to each attribute.

Regarding the fact that neither analyst nor the consumer (driver) himself or herself is not able to reliably identify all attributes influencing the value of perceived utility, the utility function is complemented with an error term that aggregates all factors unrecognized by the analyst. General form of the utility function has the form

$$U_i = V_i + \varepsilon_i,$$

[1] Transport Research Center, Lisenska 33a, 636 00 Brno, Czech Republic, petr.senk@cdv.cz.
[2] Transport Research Center, Lisenska 33a, 636 00 Brno, Czech Republic, stanislav.biler@cdv.cz.

where U_i represents utility of option i; V_i represents factors influencing decision making that are known to the analyst; and ε_i represents all remaining "unknown" factors.

The theory of rational choice suggests that consumer (car driver) who faces alternatives i and j with utilities $U_i > U_j$ always choses alternative i. However, if the utility function consists of the unknown component ε_i, as it is also in our case, the result of choice becomes more difficult to predict. The presence of random component in the utility function requires that the choice behavior has to be viewed as a stochastic process, in which given consumer (driver) choses the alternative i with probability $Prob_i$ that complies equation

$$Prob_i = Prob(U_i > U_j).$$

The decomposition of utility function to partial components leads to the equation

$$Prob_i = Prob[(V_i + \varepsilon_i) > (V_j + \varepsilon_j)]$$

that is due to the presence of random components called the rule of random utility maximization. Further adjustment leads to the equation

$$Prob_i = Prob[(\varepsilon_j - \varepsilon_i) > (V_i - V_j)]$$

that shows that the probability of choosing alternative i is equal to the probability that the difference between unknown sources of utility of alternatives j and i is higher than the difference between known sources of utility of alternatives i and j.

Further it may be shown that if random components ε_i and ε_j are drawn from type I extreme value distribution with probability function

$$Prob(\varepsilon_j \leq \varepsilon) = e^{-e^{-\varepsilon}},$$

after certain arrangements made under certain conditions (for details see e.g.[4]) it is possible to derive the alternative i is chosen over alternative j with probability

$$Prob_i = \frac{e^{V_i}}{e^{V_i} + e^{V_j}}.$$

The equation above is a specific form of discrete choice model for two alternatives – binary logit model.

2.2 Implementation of binary logit model in willingness-to-pay method

The method of willingness-to-pay consists in the estimation of maximum sum of money people are willing to sacrifice in order to gain certain merit, product or, conversely, eliminate negative consequences of their choice. In simple linear models, the willingness-to-pay may be derived as the proportion of parameter estimates related to one non-monetary attribute (in our case *time*) and one monetary attribute (in our case *travel costs*). Linearity of utility functions in discrete choice models thus enables implementation of such kind of models in the willingness-to-pay method.

In our specific case, the value of travel time during trips to leisure activities is formulated as the maximum amount of money that are people willing to sacrifice to save one unit of time, provided that all other trip related attributes remain constant. In our study we consider two hypothetical alternatives that differ on following attributes:

- Fuel related costs X_{costs} [in CZK];
- Travel time X_{time} [in *hours*];

The utility function of alternatives i and j has the form:

$$V_i = \beta_{costs} X_{costs} + \beta_{time} X_{time}$$

and

$$V_j = \beta_{costs} X_{costs} + \beta_{time} X_{time},$$

where β_{costs} and β_{time} represent respective parameters that are going to be estimated from empirical data.

Regarding the survey design (described in section 3), in which are both alternatives depicted only using values of attributes and abstract names (e.g. route G), the utility functions are free of alternative specific constants.

Parameters β_{costs} and β_{time} of the conditional logit model were estimated by the maximum-likelihood method using the *survival* package of statistical software R [5], namely the function *clogit()*. The final value of travel time savings $VTTS_{leisure}$ was obtained by substitution of estimated values of β_{costs} and β_{time} to the formula

$$VTTS_{leisure} = \frac{\beta_{time}}{\beta_{costs}} \cdot 60 \quad [in\ CZK/h].$$

3 Data

3.1 Stated preference survey

The aim of stated preference survey is to estimate weights that consumers assign to particular attributes of available alternatives. Respondents of the survey are introduced into hypothetical scenario, provided with hypothetical choice alternatives that differ on values of particular attributes and asked to choose one alternative based on their preferences.

The survey design consisted in the specification of hypothetical situation with respect to the objective of the study, i.e. estimation of value of travel time in case of leisure trips. In the study we focused on the most common type of leisure trips – weekend (two to three days) leisure trips. In the following text we will describe in details the survey design.

In the stated preference experiment we assumed several hypothetical destinations (scenarios) that can be accessed by two alternative routes i a j that differ in values of attributes X_{costs} and X_{time}. In order to reduce alternative specific bias, routes were labeled by randomly generated letters (e.g. route D and route B). Further, range of plausible values of attributes were calculated based on the assumption that one way trip in case of weekend leisure travel takes about 1 hour 40 minutes, which in case of average speed 70 km/hour correspond to 117 km. Adequacy of the assumption was later confirmed by outcomes of survey, in which respondents provided answer to the question "In case of weekend leisure trip by car, what is the average distance traveled (return trip in km)?". The average distance provided by respondents was 253 km (return trip). Travel time values were set within $\mp$30 minutes range around the average travel time, what corresponds to the average travel speed between ranging from 44 km/h to 119 km/h. The setting of travel costs was based on the average fuel consumption 8 liter per 100 km and fuel price 35 CZK per liter, what leads to the average one way travel costs approximately 330 CZK, 270 CZK for the shortest route and 390 CZK for the longest route.

With respect to the fact that variables X_{costs} and X_{time} are continuous, it was necessary to select limited number of values so that we would be able to cover in sufficient detail maximal range of travel time values, while keeping the number of decision problems as low as possible. The range of travel time values was set within the range from 10 CZK/h to 360 CZK/h, with respect to literature [3] that recommends values covering the range of 25% to 50% of average hourly wage (143 CZK/h in second quarter of 2011 [2]). Finally fractional factorial design was used in order to generate minimum number of value levels. Table 1 shows selected values of attributes.

X_{time}	X_{costs}
1 h 10 min	270 CZK
1 h 35 min	310 CZK
1 h 45 min	350 CZK
2 h 10 min	390 CZK

Table 1 Attribute values used in stated preference survey

Each respondent of the survey was provided with 9 decision problems that according to fractional factorial design systematically combined values in Table 1 and asked to choose his/her preferred route. Table 2 shows an example of card with the decision problem.

Scenario 1	**Route A**	**Route B**
Costs related to fuel consumption	70 CZK	130 CZK
Travel time	1h 45min	1h 35min

Table 2 Example of card with decision problem used in the survey

3.2 Dataset

The survey sample was selected from the population of residents in the Czech Republic using stratified random sampling based on the register of addresses. 557 out of the total of 835 selected respondents agreed to participate

in the survey, what corresponds to 67% response rate. Sample representativeness was confirmed by the comparison of socio-demographical characteristics of the sample with characteristics of the population.

Survey was executed using CAPI method. Each respondent was asked for his/her socio-demographic characteristics and basic characteristics of travel behavior in the context of leisure travel. Respondents who do not use car for leisure travel were excluded from the sample. Finally, respondents were introduced to the context of the decision problem and provided with nine decision scenarios. The final dataset consists of 2502 choices from 278 respondents.

4 Results

Table 3 shows final values of parameters β_{costs} and β_{time} estimated on the above described sample.

Attribute	Parameter estimate β	SE	z-score	Pr (>|z|)
X_{costs}	-0,041527	0,002048	-20,27	< 0,001
X_{time}	-0,038672	0,002428	-15,93	< 0,001

Table 3 Estimated parameters of conditional logit model

Negative values of parameters correctly suggest that higher travel costs, respectively higher travel time, have negative impact on utility. Low p-values suggest that estimated values of both parameters are significantly different from zero.

The value of travel time savings in case of weekend leisure trips is

$$VTTS_{leisure} = \frac{\beta_{time}}{\beta_{costs}} \cdot 60 = \frac{-0{,}038672}{-0{,}041527} \cdot 60 = 55{,}87 \quad [CZK/h].$$

Estimated value of travel time savings is within the range referred in literature [3], i.e. 25% - 50% of average hourly wage in the Czech Republic.

5 Discussion

This study presented estimation of value of travel time savings in case of weekend leisure travel using willingness-to-pay method that combined stated preference survey and conditional logit model. It was shown that estimated values of travel time saving are consistent with values found in other countries.

However, it has to be noticed that the study leaves aside several important issues, particularly due to the small sample size and limited scope of the study. Further study should focus on the influence of socio-demographic characteristics and detailed characteristics of trip (e.g. trip length or other types of leisure travel) on perceived value of travel time savings.

References

[1] Bickel, P., Friedrich, R., Burgess, A., Fagiani, P., Hunt, A., Jong, G. D., and Laird J. (eds.): *Harmonised Guidelines for Project Assessment, Deliverable 5, HEATCO - Developing Harmonised European Approaches for Transport Costing and Project Assessment*, EC 6 FP. IER, 2006, University of Stuttgart.

[2] Informační systém Ministerstva práce a sociálních věcí. http://www.mpsv.cz/files/clanky/11523/ISPV_112.pdf.

[3] Litman TA, Doherty E. *Transportation cost and benefit analysis: techniques, estimates and implications*. (2nd edition), Victoria Transport Policy Institute, Victoria, British Columbia, 2009.

[4] Maddala G S. *Limited dependent and qualitative variables in econometrics*. Cambridge University Press, Cambridge, England, 1983.

[5] R Development Core Team. *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0, URL http://www.R-project.org, 2008.

Bilateral J-Curve between Slovakia and its major trading partners

Jana Šimáková[1]

Abstract. The primary purpose of this paper is to analyse the impact of exchange rate on bilateral trade flows between Slovakia and its seven major trading partners. Devaluation or depreciation of a currency worsens the trade balance before improving it, resulting in a J-curve pattern. This paper investigates J-curve phenomenon using quarterly time series data over the period 1997:1 to 2010:4. In this paper is applied the Johansen cointegration test to analyse the long run relationship between exchange rate and trade balance. Short term effects and the related J-curve effect is explored by estimating error correction model and by assessing the impulse response function of the trade balance on the exchange rate shock. The results reveal that the trade balance between Slovakia and its trading partners does not support the J-curve phenomenon in all cases. The J-curve phenomenon was revealed in case of Hungary and partially in Czech Republic, in case of Austria S-curve pattern was found, in other cases, the coefficient estimates follow any specific pattern in response to currency depreciation.

Keywords: J-curve, trade balance, exchange rate, international trade
JEL Classification: F10, F14, F31
AMS Classification: 62M10

1 Introduction

Slovakia, as a small open economy, depends on the rest of the world and the level of interdependence has increased in last two decades. Domestic market is small to support large scale demand and inevitably depends on imports from other countries to supply a part of domestic consumption. On the other hand, there is a huge amount of exported slovak production to the other countries. This fact makes Slovak economy vulnerable to any adverse changes in other economies. One of the most important factors which influence international trade development is exchange rates.

Discussion about exchange rates and its impact on international trade flows begun in 1973, after the Bretton-Wood system collapse. This transformation period has brought obvious volatility and uncertainty. Despite the many research dealing with this interconnection, results are not always clear. From theoretical point of view, if the Marshall-Lerner condition holds, an improvement in the trade balance would occur [14]. On the empirical level, it is still open issue. There is some support in theory for the pattern known as J-curve effect. It means that at the beginning trade balance detoriates before it subsequently improves. There are numerous empirical studies exploring this issue, some of these are mentioned in following section. Their findings are mixed and depend on used data and methodology.

The aim of this paper is to explore whether exchange rate depreciation improves Slovakia´s trade balance and vice versa. Data used in this study covers period from 1997 to 2010.The methodology used while exploring the long run relationship between exchange rate and trade balance in this paper is cointegration analysis. In this paper is used Johansen cointegration test [10]. Short term effects and the related J-curve effect is explored by estimating error correction model and by assesing the impulse response of the trade balance on the exchange rate shock.

This paper continues as follows. In the next section theoretical framework and some of the previous research are reviewed and reported. Section 3 offers model specification. The presence of a J-curve phenomenon is explored in section 4, employing an vector error correction model and impuls responses. Setion 5 concludes the paper.

2 Theoretical framework and literature review

The theoretical basis of the J-curve comes from Marshall and Lerner. The Marshall-Lerner condition states that the sum of export and import demand elasticity has to be at least one and that the currency devaluation will have a positive impact on trade balance. As a devaluation of the currency means a reduction in the price of exports,

[1] Silesian University in Opava, Department of Finance, Univerzitní náměstí 1934/3, 733 40 Karviná, simakova@opf.slu.cz.

quantity demanded for these will increase. At the same time, price of imports will rise and their quantity demanded will decrease.

Empirical examination of the Marshall-Lerner condition has a long history with very different views. It has been found that goods tend to be inelastic in the short term, as it takes time to change consuming patterns [4]. Thus, the Marshall–Lerner condition is not met, and devaluation is likely to worsen the trade balance initially. In the long term, consumers will adjust to the new prices, and trade balance will improve. This effect is called J-curve phenomenon.

As to the short run effect and J-curve effect it was first advanced by Magee [13], after the fact that short run deterioration and long run improvement after currency depreciation resemble the letter "J" as seen in Figure 1. After that a large number of empirical studies exploring this problem has appeared. They explore long run impact of exchange rate on trade balance and whether J-curve effect is present.

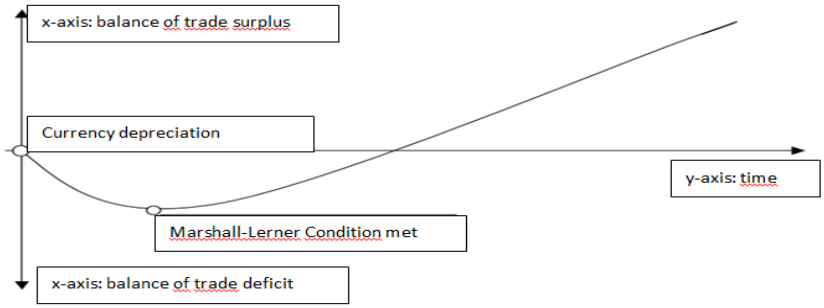


Figure 1 J-curve pattern [6]

Junz and Rhomberg [11] has attributed the J-curve phenomenon to lags in the recognition of exchange rate changes, in the decision to changes of real variables, in delivery time, in the replacement of inventories and materials, and in production. Krueger [12] has explained the phenomenon by the fact that at the time an exchange rate change occurs, goods already in transit and under contract have been purchased, and the completion of those transactions dominates the short term change in the trade balance. Therefore, exchange rate change first deteriorates the trade balance, but as the elasticities increase, it improves the trade balance. This phenomenon is not always aplicable in each country.

Bahmani-Oskooee and Kutan [3], by using monthly data over the January 1990 and June 2005 period from eleven European emerging economies (including Slovakia), after applying ARDL cointegration approach and corresponding error correction model, found empirical support for the J-curve effect in three countries: Bulgaria, Croatia and Russia - short run deterioration combined with long-run improvement was revealed. It was solved on the bilateral industry level trade balanace between two countries. Stučka [16] showed the existence of the J-curve also in Croatia. The ARDL cointegration approach used quarterly data. Hsing [9] examined the J-curve for the bilateral trade between Croatia, the Czech Republic, Hungary, Poland, Slovakia, or Slovenia and the USA. This paper found that the J-curve is not empirically confirmed for any of these six countries. In case of Slovakia, after a shock to real depreciation, the trade balance detoriates.

Using generalized impulse response functions, Hacker and Hatemi [8] tested the trade J-curve for three transitional central European countries – the Czech Republic, Hungary, and Poland – in their bilateral trade with respect to Germany. Their findings suggest that for each country there are some characteristics associated with a J-curve effect: after a depreciation trade balance decreases within a few months and then rises to a long run equilibrium value higher than the initial one. Ferreira-Lopes and Neves Sequeira [15] assessed the existence of a S-Curve pattern in ten Central and Eastern European Countries for the relation between the trade balance and the terms of trade. Empirical results support the existence of this curve for Slovenia, Czech Republic, Hungary. In the case of Slovakia the pattern is weaker than in the mentioned countries but is stills prevail.

3 Model specification

The consensus among all recent studies is that the trade balance should depend on a measure of domestic income, a measure of foreign income and the real exchange rate. In order to detect the long term co- movement among the variables, the cointegration procedure developed by Johansen [10] is used. Thus, following Bahmani-Oskooee and Kutan [3] and many other studies, equation (1) is adopted:

$$\ln TB_t = \alpha + \beta \ln Y_{d,t} + \gamma \ln Y_{f,t} + \lambda \ln REX_t + \varepsilon_t \qquad (1)$$

In equation (1) *TB* is a measure of trade balance defined as the ratio of Slovakia's exports to country *f* to her imports from country *f* ; Y_d is a measure of Slovakia's real income set in index form to make it unit free [2]; Y_f is the index of real income in trading partner *f* and *REX* is the real bilateral exchange rate between Slovakia and trading partner *f* defined in a way that an increase reflects a real depreciation of Slovak currency to currency of trading partner *f* . Estimate of β could be positive or negative. Usually an increase in domestic income leads to higher imports, what means positive estimate for β. However, if the increase in domestic income is due to an increase in production of import substitute goods, imports could actually decline, what means a negative β [1]. According to this, estimate of γ could also be positive or negative. REX is defined in a way that an increase reflects real depreciation of domestic currency. If real depreciation of domestic currency improves her trade balance with partner *f*, an estimate of λ is expected to be positive [4].

The above trade balance model represents the long run relationships between the trade balance and its determinants. To test the J-curve phenomenon (short term relationship), a short term dynamics must be incorporated into the long run model. According to Hsing [9] it is modified to error correction modeling format. In this case, the error correction model is as follows:

$$\Delta \ln TB_t = \alpha + \sum_{k=1}^{K} \omega_k \Delta \ln TB_{t-k} + \sum_{k=1}^{K} \beta_k \Delta \ln Y_{d,t-k} + \sum_{k=1}^{K} \gamma_k \Delta \ln Y_{f,t-k} + \sum_{k=1}^{K} \lambda_k \Delta \ln REX_{t-k} \tag{2}$$

4 The results

This section reports the estimates of J-curve for Slovakia and her major trading partners. The vector error correction model (2) is estimated for seven major Slovakia´s trading partners: Austria, Czech Republic, France, Germany, Hungary, Italy and Poland, using quarterly data over the period 1997 to 2010. The selection of sample countries is based on share of total foreign trade turnover of Slovakia. Data are obtained from OECDiLibrary.

Before conducting other tests, time series are modified by logarithmic transformation. This helps reduce skewness and heteroscedasticity and to stabilize variability. The stability of regressors is needed in initial testing. Before estimating the cointegration parameters, the order of integration for each time series should be examined. Integration is determined using the augmented Dickey-Fuller test recommended by Engle and Granger [7]. ADF test for each individual time series confirmed the presence of unit roots and for all variables was found the first-difference stationarity. According to Blake and Fomby [5], non-stationarity on the first level I (1) is the basic precondition of cointegration between variables. In applying the Johansen procedure, it is needed to specify the number of lags in each cointegration equation. Optimal lags in cointegrated time series are based on Akaike Information Criterion and Schwarz Bayesiani Criterion. Results of cointegration procedure are in Table 1.

	Lag	Trace Statistic	Critical Value at 5%	Max-Eigen Statistics	Critical Value at 5%	No. of CE
Austria	5	35.09788	29.79707	21.91651	21.13162	r=2
Czech Republic	2	65.52314	47.85613	34.21242	27.58434	r=1
France	5	76.95363	47.85613	32.96911	27.58434	r=1
Germany	5	32.67442	29.79707	23.21361	21.13162	r=2
Hungary	1	66.70999	47.85613	29.21835	27.58434	r=1
Italy	1	64.37987	47.85613	32.60939	27.58434	r=1
Poland	1	59.26424	47.85613	29.95219	27.58434	r=1
Note: 'r' refers to the number of cointegrating vector which are significant under both tests. The critical values are from MacKinnon-Haug-Michelis (1999).						

Table 1 The results of cointegration procedure

According to Table 1, the null hypothesis which means no cointegration can be rejected and it can be concluded that there is evidence for cointegration among these variables. Hence, these variables should be retained in the model.

	Austria	Czech Republic	France	Germany	Hungary	Italy	Poland
ΔlnREX	-0.568943	-0.721994	2.175522	1.655021	-1.765323	7.613189	-0.351836
	(1.10363)	(0.49113)	(0.80396)	(0.74296)	(1.26944)	(2.18217)	(0.06152)
ΔlnREX$_{t-1}$	-0.272752	-0.521229	0.032102	0.122986	-0.054983	0.314381	0.136150
	(0.38692)	(0.29895)	(0.38026)	(0.07301)	(0.50313)	(0.70176)	(0.12491)
ΔlnREX$_{t-2}$	-0.246366	0.007815	-0.104801	0.093252			
	(0.41014)	(0.30450)	(0.43975)	(0.07270)			
ΔlnREX$_{t-3}$	-0.100033		-0.580613	0.093958			
	(0.38198)		(0.47113)	(0.07117)			
ΔlnREX$_{t-4}$	0.266375		-0.792028	0.067719			
	(0.38900)		(0.51141)	(0.07056)			
ΔlnREX$_{t-5}$	0.758932		-0.101349	0.090621			
	(0.36440)		(0.40890)	(0.06004)			
EC	-0.074777	-0.028764	-0.078975	-0.020244	0.016175	0.020754	2.977690
	(0.02351)	(0.02401)	(0.02908)	(0.11428)	(0.00824)	(0.00571)	(0.48252)
Note: Figures in the parentheses are absolute values of t-statistics.							

Table 2 Estimated coefficients of the real Exchange rate and error correction term

As indicated before, the short run effects of depreciation are reflected in the coefficient estimates obtained for the lagged value of the first differenced exchange rate variable. The J-Curve phenomenon should be supported by negative coefficients followed by positive ones. To construct J-curve, impulse response function is used. The result can be seen in Figures 2 to 8.

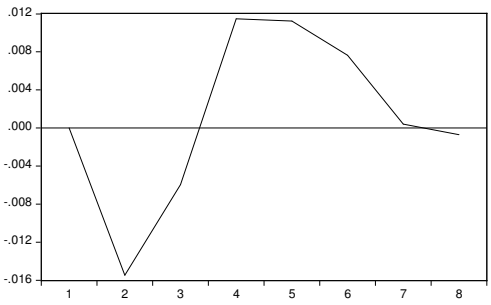


Figure 2 Austria: Response of LNTB to LNREX impulse*

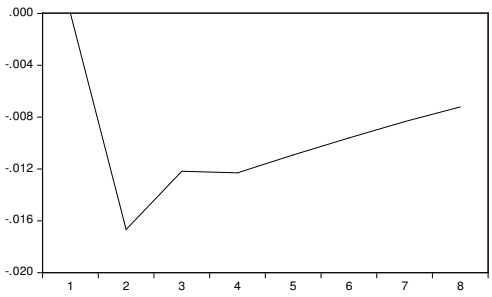


Figure 3 Czech Republic: Response of LNTB to LNREX impulse*

* X axis shows trade balance deficit or surplus
Y axis shows time

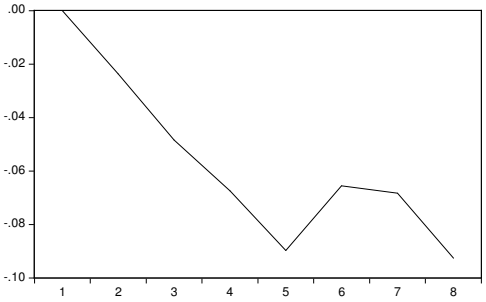


Figure 4 France: Response of LNTB to LNREX impulse*

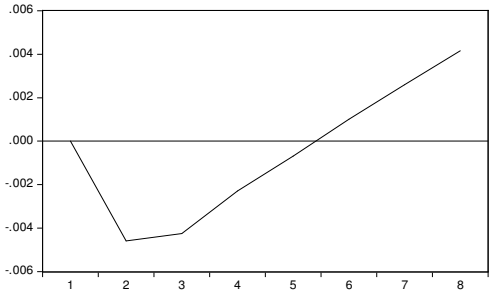


Figure 5 Hungary: Response of LNTB to LNREX impulse*

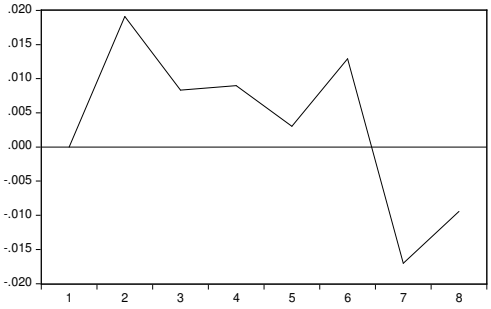


Figure 6 Germany: Response of LNTB to LNREX im-

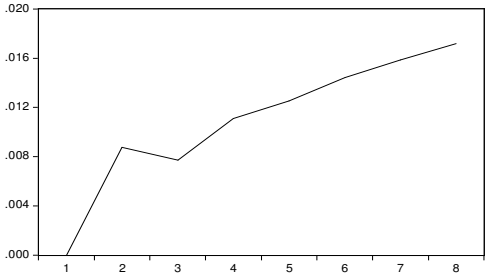
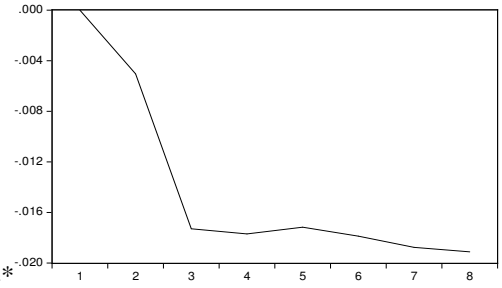


Figure 7 Italy: Response of LNTB to LNREX impulse*



pulse*

Figure 8 Poland: Response of LNTB to LNREX impulse*

From Table 2 and Figures 2 to 8 results, that J-curve phenomenon was revealed in case of Hungary and partially in Czech Republic, in case of Austria S-curve pattern was found, in other cases, the coefficient estimates follow any specific pattern.

5 Conclusion

To explore the J-curve effect, disaggregated bilateral data from Slovakia and her seven largest trading partners are used. The short run and the long run response of the trade balance to currency depreciation was investigated. The methodology was based on Johansen cointegration procedure, vector error correction model and impulse response function.

Estimation of cointegration procedure revealed a long term relationship between Slovakia´s trade balance, real bilateral exchange rate, Slovakia´s gross domestic product and gross domestic product of each trading partner. J-curve phenomenon was revealed in case of Hungary and partially in Czech Republic, in case of Austria S-curve pattern was found, in other cases, the coefficient estimates follow any specific pattern.

Acknowledgement

Research was supported by the student grant competition SGS/25/2010: Financial Integration in the EU and its Impacts on Corporate Sector.

References

[1] Bahmani-Oskoee, M.: Determinants of international trade flows : The Case of Developing Countries. *Journal of Development Economics* **20** (1986), 107-123.

[2] Bahmani-Oskoee, M.: Is there a long-run relation between the trade balance and the real effective exchange rate of LDCs*? Economics Letters* **36** (1991), 403-407.

[3] Bahmani-Oskoee, M. and Kutan, A.: The J-curve in the emerging economies of Eastern Europe. Applied Economics. *Taylor and Francis Journals* **41** (2009), 2523 -2532.

[4] Bahmani-Oskoee, M. and Ratha, A.: The J-Curve: a Literature Review. Applied Economics. *Taylor and Francis Journals* **36** (2004), 1377-1398.

[5] Balke, N. and Fomby, T.: Treshold cointegration. *International Economic Review* **38** (1997), 627-645.

[6] Clarke, A. and Kulkarni, G.: Testing the Application of Heckscher-Ohlin Theorem to Contemporary Trade Between Malaysia and Singapore. *Journal of Emerging Knowledge on Emerging Markets* **1** (2009).

[7] Engle, R. and Granger, C.: Co-integration and Error Correction: Representation, Estimation, and Testing. *Econometrica* **55** (1987), 251-276.

[8] Hacker, S.R. and Hatemi, A.J.: The effect of exchange rate changes on trade balances in the short and long run. *The Economics of Transition, The European Bank for Reconstruction and Development* **12** (2004), 777-799.

[9] Hsing, Y.: Responses of Output to Declining Stock Values and Real Depreciation in Lituania. *International Economics* **62** (2009), 429-437.

[10] Johansen, S.: *Likelihood-Based Interference in Cointegrated Vector Autoregressive Models.* Oxford University Press, Oxford, 1997.

[11] Junz, H. and Rhoemberg, R. R.: Price Competitiveness in Export Trade among Industrial Countries. *Papers and Proceedings of American Economic Review* **63** (1973), 412-418.

[12] Krueger, A.: The Factor Proportions Explanation of Trade, Distortions, and Employment. In: *Trade and Employment in Developing Countries.* National Bureau of Economic Research, 1983, 55-85.

[13] Magee, S.P.: Currency Contracts, Pass Through and Devaluation. *Brooking Papers on Economic Activity*, **1** (1973), 303-325.

[14] Petrović,P. and Gligorić, M.: Exchange Rate and Trade Balance: J-curve Effect. *Panoeconomicus, Savez ekonomista Vojvodine,* **57** (2010), 23-41.

[15] Sequiera, T.N. and Lopes, A.: On the welfare properties of the Lucas and Romer endogenous growth models. *Economics Bulletin* **30** (2010), 2143-2150.

[16] Stučka, T.: The Impact of Exchange rate Changes on the Trade Balance in Croatia. *CNB: Croatian National Bank Working Paper Series no. W – 11,* 2003.

Faster convergence for estimates of parameters of Gompertz-Makeham function using available methods in solver MS Excel 2010

Ondřej Šimpach[1]

Abstract. In basic applications of specific mortality rates we use moving averages of different lengths and different set of weights. Given that the weighted MAs aren't enough, it's useful to choose one of the sophisticated functions. One of these functions is Gompertz–Makeham and also modified G–M, which are both of exponential type. The initial estimates of parameters can be set up manually and using by other software it's possible to refine it. Given that the optimization SW is relatively expensive, it's useful MS Excel Solver, which is able to solve simple LP and NLP tasks within a reasonable time. Even in v. 2007 there was available basic method to solve simple LP and NLP only, and the additional package was necessary to purchase externally. Since v. 2010 is incorporated in standard package GRG–NL and Evolutionary method, which solve more difficult NLP in less time than before. A subsequent study will provide a comparison of computational time needed to obtain accurate estimates of parameters of functions in solvers v. 2007 and 2010, based on real data of selected EU countries. If is possible to refine the initial parameters of this function, then using solver 2010 and in less time.

Keywords: Gompertz-Makeham, MS Excel 2010, specific death rates, convergence.

JEL Classification: C61, C63
AMS Classification: 90C30

1 Introduction

Optimization approaches are also important in demographics, because for special types of applications it is necessary to estimate and refine the parameters of specific functions. One of these function is Gompertz-Makeham function for balancing specific death rates (see Gompertz [6], Makeham [10]), which is used in the construction of complete life tables for commercial use (see e.g. the projection of the human capital of the Czech Republic and its regions to 2050 by Fiala et al. [5] or the development of life expectancy in the Czech Republic and EU countries by Langhamrová Jana and Langhamrová Jitka [9]). This function is of exponential type and therefore is effective use a special procedure for its solution. The initial estimates of its parameters can be calculated from empirical data (see Cipra [1]) and using the optimization software is also possible to improve these estimates. Because the optimization software is expensive and in comparison with other applications of operations research, the balancing of specific death rates is relatively simple application, it is not effective to invest in making professional software, but it is better to use available solutions in a readily available application, e.g. MS Office.

The solver in MS Excel 2007 was equipped with only basic module for solving linear programming and simple nonlinear programming. Additional modules for solving more difficult problems were only available after the external purchase. This has raised the question of whether is better to buy optimization software directly than a supplement to Microsoft Office suite. With the coming of version of MS Excel 2010 the modules GRG Nonlinear and Evolutionary have been incorporated into the solver, which can be used to solve more difficult nonlinear problems. These modules are available from the 2010 release in the basic installation of the base price, so there is not necessary to re-buy it externally.

2 Modeling of specific death rates

In demographic analysis we use the modeling of specific death rates particularly for the purpose of life insurance (see Cipra [2] or Fiala [3]). The life tables describe the dependence of mortality on the age. The basic characteristics for their calculation are therefore specific death rates. In the case of life tables for one calendar year t (see Fiala [4]) are specific death rates calculated by the formula

[1] University of Economics Prague, Department of Demography, Winston Churchill sq. 4, 130 67 Prague, ondrej.simpach@vse.cz

$$m_x = \frac{M_{t,x}}{\frac{S_{t,x} + S_{t+1,x}}{2}}, \tag{1}$$

where $M_{t,x}$ is the number of death in the age of x in year t, $S_{t,x}$ is the initial state of x-years old person in year t. The development of specific death rates for the Czech Republic in 2000 is shown for illustration in Figure 1 (data source: Eurostat, own construction) and the history of mortality in selected European countries can be found in [8]).

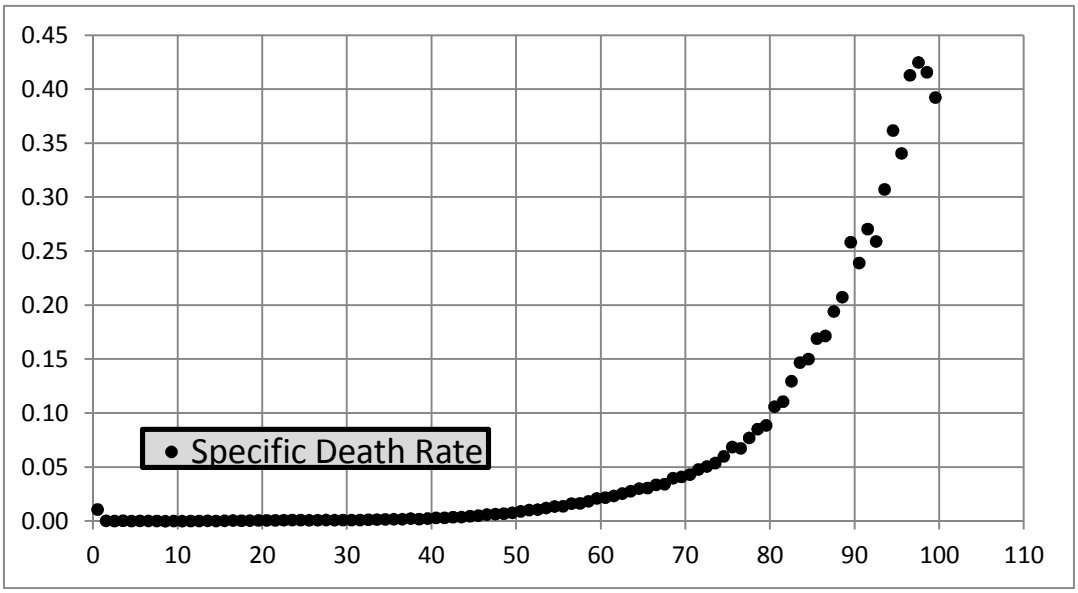


Figure 1 The specific death rates for the Czech Republic in 2000.

Death rates are developed up to particular age x (usually $x = 60$) by previously expected system. For different types of populations are these values more or less scattered in higher age, and for the purpose of further analysis is necessary to balance them. This balancing is performed using either a weighted moving averages (simple method), or with sophisticated balancing function (complex method). High quality outputs provides Gompertz-Makeham balancing function for specific death rates (see Koschin [7]), defined as

$$\tilde{m}_x^{(GM)} = a + b \cdot c^{x+\frac{1}{2}}, \tag{2}$$

for which we set up the beginning of the first age interval $x_0 = 60$ and the length of interval $k = 8$. The initial estimations of parameters of Gompertz-Makeham function is selected according to the recommendations (see Fiala [4]) the following algorithm:

Algorithm 1. Calculate the empirical sums of specific death rates in the various intervals as

$$G_1 = \sum_{x=60}^{67} m_x, \tag{3}$$

$$G_2 = \sum_{x=68}^{75} m_x \tag{4}$$

and

$$G_3 = \sum_{x=76}^{83} m_x. \tag{5}$$

Next calculate c^k and c (where $k = 8$) as

$$c^8 = \frac{G_3 - G_2}{G_2 - G_1}, \tag{6}$$

$$c = \sqrt[8]{c^8}\ . \tag{7}$$

Calculate the value of the auxiliary expression

$$K_c = c^{60,5} \cdot (1 + c + ... + c^7) = c^{60,5} \cdot \frac{c^8 - 1}{c - 1}, \tag{8}$$

and then

$$b = \frac{G_2 - G_1}{K_c \cdot (c^8 - 1)}, \tag{9}$$

$$a = \frac{G_1 - b \cdot K_c}{8}. \tag{10}$$

The values of Gompertz-Makeham function (2) we calculated for the values of age from 60 to 83 years. As the values of the parameters *a*, *b* and *c* we use the values, that will refine the least squares method. For the age interval $x \in \langle 83;\ 110 \rangle$ we use instead of the original Gompertz-Makeham function its modification (see Koschin [7]), defined as

$$\tilde{m}_x^{(GM)'} = a + b \cdot c^{83 + \left[\ln(d \cdot (x + \frac{1}{2} - 83) + 1)\right] / d}. \tag{11}$$

As an initial estimate of the parameter *d* we choose for example *d* = 0.02. Next, for age 83 we overwrite the value given by Gompertz-Makeham function (2) with modified Gompertz-Makeham function given by formula (11). The value of *d* will be optimized using least squares method. Weighted squares of deviations, which are defined as

$$\frac{S_{t,x} + S_{t+1,x}}{2 \cdot m_x \cdot (1 - m_x)} \cdot (m_x - \tilde{m}_x^{(GM)})^2, \tag{12}$$

we calculate for all values of *x* from age 60 to *y*, where *y* is the maximum age for which we have empirical death rate with non-zero value. Weighted sums of squared deviations we calculate separately for the age in range 60–82 years (the sum of S_1), where we use the original (unmodified) function, and particularly for the age in range of 83–*y* (sum of S_2), where we use a modified function. Optimization procedure is performed with the "solver", implemented in MS Excel. Optimization is performed twice: first separately for sum of squares S_1 (when we change the parameters *a*, *b* and *c*) and separately for the second sum of squares S_2 when we change only the parameter *d*. Because we used the empirical values of specific death rates to estimate the weights, we do not have to insert any limiting conditions.

3 Results and comparison of computational times

For each country of the European Union, the procedure was repeated ten times with the concrete solver and the computing times were measured. Ten times the measurements were taken in a standard module in solver of MS Excel 2007, ten times the measurements were taken in the module GRG Nonlinear in solver in MS Excel 2010, and finally ten times the measurements were taken in the module Evolutionary in solver in MS Excel 2010. From these times there was calculated the average time, that has been necessary for the calculation of the corrected estimations of parameters of Gompertz-Makeham function and modified Gompertz-Makeham function. The used CPU was Intel ® Core ™ i5-2430 2.40 GHz running at 64-bit platform. An overview of these computational times is summarized in Table 1.

Country	Standard Solver MS Excel 2007	GRG Nonlinear MS Excel 2010	EVOLUTIONARY MS Excel 2010
Belgium	5.048	2.793	1.862
Bulgaria	10.648	2.793	1.893
Czech Republic	6.655	5.586	5.048
Denmark	13.31	8.379	3.155
Germany	18.634	12.103	8.834
Estonia	11.979	5.586	5.048
Ireland	13.31	12.103	11.172
Greece	11.979	8.379	6.31
Spain	11.979	7.448	3.786
France	10.241	6.941	5.048
Italy	10.241	5.048	6.31
Cyprus	11.172	8.834	4.417
Latvia	11.979	8.379	8.203
Lithuania	17.303	13.034	8.834
Luxembourg	9.317	4.655	1.262
Hungary	6.941	4.655	3.993
Malta	18.634	13.034	6.31
Netherlands	18.634	11.172	4.417
Austria	8.203	5.586	4.655
Poland	6.655	6.517	5.572
Portugal	10.648	6.941	3.724
Romania	18.634	2.793	1.262
Slovenia	11.979	6.517	3.155
Slovakia	17.303	9.317	6.517
Finland	11.172	6.655	3.155
Sweden	4.655	3.724	1.262
United Kingdom	13.31	9.31	4.417

Table 1 Computing times for EU countries (in seconds).

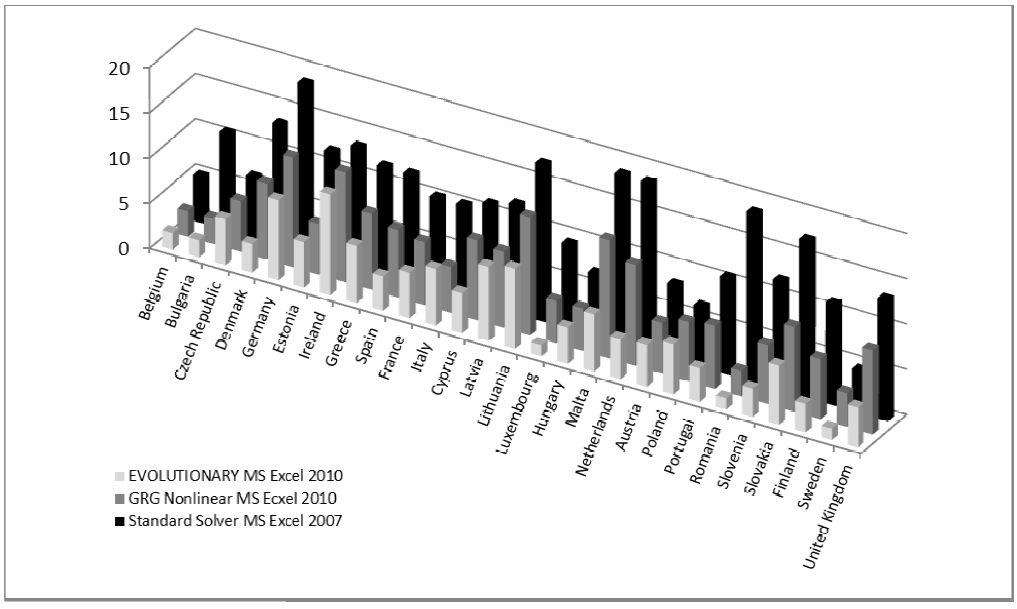


Figure 2 Computing times for EU countries (in seconds).

The values in Table 1 are presented graphically in Figure 2. It is clear that in all cases the calculation module GRG Nonlinear provides solutions in less computing time than the standard module, implemented in MS Excel 2007. In most cases, the Evolutionary module provides a solution before GRG Nonlinear module, but not always. In Figure 3, individual EU countries are organized in descending order of how long it took to refine the parameters of estimates of its Gompertz-Makeham function. The solution provided by individual modules, of course, depends on the input data and on other circumstances. However, we can confirm, that the modules GRG Nonlinear and Evolutionary, implemented in MS Excel 2010, provide a solution sooner than in older versions of MS Excel, what can **bring benefits for specific groups of users and analysts.** It is also important to note that there is probably no relationship between the length of computing time and the size of the considered country. Even if this relationship will be found, this correlation will be certainly misleading. In the case that this study would be performed for another calendar year with other statistics of deaths, the length of computing time would be certainly changed for some countries and the ranking of countries would then also be changed. Probably would not change the order of considered modules.

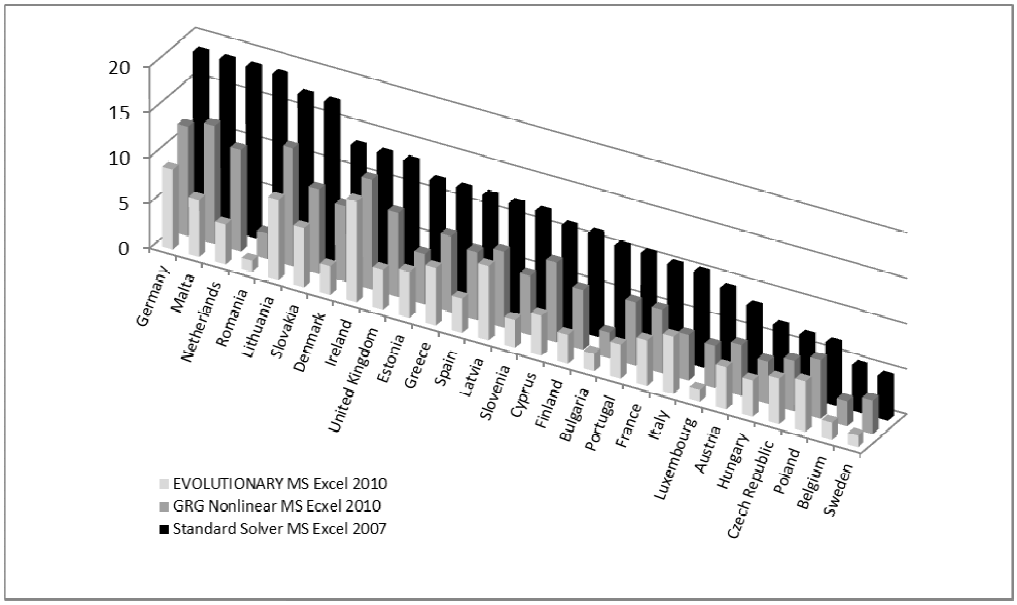


Figure 3 Computing times for EU countries (in seconds, organized in descending order).

References

[1] Cipra, T.: *Matematické metody demografie a pojištění*. SNTL, Praha 1990.
[2] Cipra, T.: Life tables for Czech pension funds. *Insurance: "Mathematics & Economics"* 32 (2003), 176–177.
[3] 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, roč. 7, č. 4, s. 35–44.
[4] Fiala, T.: *Demografické výpočty v tabulkovém procesoru.* 1. vyd., Praha : Oeconomica, 2002.
[5] Fiala, T., Langhamrová, J., Průša, L.: Projection of the Human Capital of the Czech Republic and its Regions to 2050. *Demografie*, 2011, roč. 53, č. 4, s. 304–320.
[6] 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.
[7] Koschin, F.: *Aktuárská demografie.* 1. vyd., Praha : VŠE, 1993.
[8] Langhamrová, J., Fiala, T.: Přehled úmrtnosti v ČR a ve vybraných evropských zemích v letech 1950-1992. *Statistika*, 1996, roč. 33, č. 5, s. 200–209.
[9] Langhamrová, Jana, Langhamrová, Jitka.: *Vývoj střední délky života v České republice a v zemích EU.* Bratislava : SAV; ČAV, 2010. 36–44 s.
[10] Makeham, W. M. "On the Law of Mortality and the Construction of Annuity Tables". *J. Inst. Actuaries and Assur. Mag.* **8** (1860): 301–310.

Exchange rate prediction: a wavelet-neural approach

Irena Šindelářová[1]

Abstract. The paper deals with the use of artificial neural networks (ANN) for predictions of economic time series. First, we revise the basic existing ANN architectures for time series forecasting and describe their application on CZK/EUR exchange rate prediction. Next, we introduce a hybrid version of the ANN that builds upon the same network strategy but tries to enhance the prediction accuracy by first decomposing the signal into individual frequency bands and then training the perceptron parameters on each frequency band separately. The time-frequency signal analysis is obtained using discrete wavelet transform. Different signal segments in various levels of decomposition are reconstructed separately. The resulting signals (in time domain again) are samples for ANNs as above. Finally, we compare the accuracy of the hybrid approach to that of the traditional ANN setting for the case of CZK/EUR exchange rate data.

Keywords: artificial neural network, wavelet transform, digital signal processing, CZK/EUR exchange rate prediction

JEL Classification: C45
AMS Classification: 65T60, 82C32

1 Introduction

In the paper the time series of CZK/EUR exchange rate is predicted. The daily data from the year 2001 to 2012 is considered. Time series forecasting is based on ANN with one hidden layer, discussed in detail in Section 2. The discrete wavelet transform (DWT) is used as a filter bank for various frequency bands obtaining. DWT is presented in Section 3. The essential idea of the hybrid wavelet-neural system presented is to forecast higher signal frequencies only. It is proposed that slow components corresponding to the low frequencies do not influence the predicted value. Only chosen components of the original signal go to the ANN of the same architecture. All computations were performed in the Matlab software.

2 Neural networks

The neural network is one of the computational models used in artificial intelligence. The pattern is the behavior of the relevant biological structures. ANN consists of formal neurons, which are interconnected and transform signals using certain transmission functions. Neuron has any number of inputs, but only one output. One of the most widely used models is the following [3]: $a = f(\mathbf{wx} + b)$. First, the vector input **x** is multiplied by the vector of synaptic weights **w**. Second the product **wx** is added to the scalar bias *b* to form the net input. Finally, the net input is an argument for the transfer function *f*, which produces the scalar output *a*. Weights express experience store to the neuron. The higher the value, the more important given input is. Bias denotes the neuron activation threshold. Depending on the type of neuron an appropriate transfer function is used (linear or sigmoid).

Multilayer perceptron

The data prediction is based on *multilayer perceptrons* (MLP) with three layers: six neurons are in the input layer, one to six neurons in the hidden layer with a log-sigmoid transfer function and one neuron in the output layer with a linear transfer function. The suggested perceptron is illustrated in Figure 1. The MLPs realize the function $x_m \approx f(x_{m-6}, ..., x_{m-1})$, where $x_1,\ldots,x_N$ denotes an original signal. The networks were trained with the momentum back-propagation algorithm, for more details see [3].

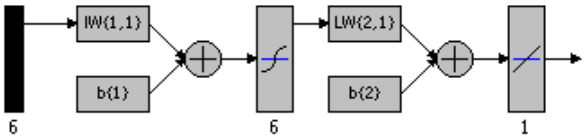


Figure 1 Multilayer feed-forward network

[1] University of Economics in Prague, Department of Econometrics, W. Churchill sq. 4, Prague, isin@vse.cz.

3 Wavelet transform

Mathematical Background

A *wavelet* is an oscillating function which can be used as a sine wave in Fourier analysis. It expands signals in terms of sinusoids having infinite energy. On the other hand a wavelet has finite energy concentrated in time around a point (Figure 2). While Fourier transform provides a given signal frequency analysis, a wavelet transform gives both time and frequency simultaneously [2].

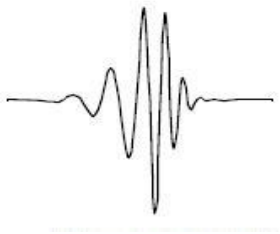


Figure 2 Example of wavelet

Wavelets are functions $w_{j,k}(t)$ with parameters of translation k a and scale j, where j, k = 1, 2, … The family of functions derived from the *initial wavelet* creates an orthonormal system:

$$w_{j,k}(t) = \frac{1}{\sqrt{j}} w\left(\frac{t-k}{j}\right)$$

The scale decreasing is referred to the dilation of a wavelet and translation is a shifting along the time axis. The *discrete wavelet transform* of a given signal f is the comparison of the signal and wavelets with varying k and j. The correlation rate of this comparison is expressed by *wavelet coefficients* ($a_{j,k}$). Therefore the wavelet expansion represents signal components with a more accurate local description than Fourier coefficients.

$$f(\mathrm{t}) = \sum \mathrm{a}_{j,k}\, 2^{j/2}\, w(2^j t - k)$$

Signal processing

Wavelet coefficients $a_{j,k}$ as the given signal representation in the system of wavelets can be organized into levels (various signal scales). This *multiresolution* representation [4] separates signal frequency information into selected number of levels of decomposition. The input signal length (N) has to be of power of 2. The number of decomposition levels can be equal to maximal n where $N = 2^n$. The described process is a signal filtering in fact. It is possible to develop most of the wavelet theory results using *filter banks* [2].

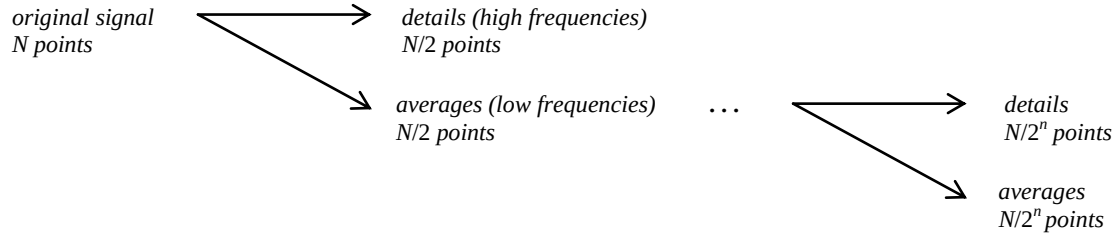


Figure 3 Principle of multiresolution

Real data processing

The time series of CZK/EUR exchange rate from October 2000 to April 2012 is considered. The number of observations is 2048, i.e. n = 11. We denote the vector of details in level k as $\mathbf{d}^k$, and averages as $\mathbf{a}^k$. Lengths of $\mathbf{d}^k$ and $\mathbf{a}^k$ are 2^{n-k}. The given signal discrete wavelet transform for n as a number of levels results in the vector:

$$\mathbf{r}^* = [\mathbf{a}^n, \mathbf{d}^n, \mathbf{d}^{n-1}, \ldots, \mathbf{d}^1].$$

The Daubechies $D4$ wavelet in the decomposition is proposed for our purposes. $D4$ wavelet function is defined by the dilation equation in the form

$$w(x) = -c_3 w(2x) + c_2 w(2x-1) - c_1 w(2x-2) - c_0 w(2x-3),$$

where $c_0 = (1+\sqrt{3})/4$, $c_1 = (3+\sqrt{3})/4$, $c_2 = (3-\sqrt{3})/4$, $c_3 = (1-\sqrt{3})/4$.

4 Hybrid wavelet-neural system

This approach is based on the ANN prediction of the signal frequency bands separately. The input data of the proposed MLP defined in Section 1 are preprocessed data by means of DWT and iDWT. The number of decomposition level is 11 and $r^* = [a^{11}, d^{11}, d_1^{10}, d_2^{10}, d_1^9, d_2^9, d_3^9, d_4^9, \ldots, d_1^1, \ldots, d_{1024}^1]$. Then the signal transformation back to the time domain is following but it will not be perfect reconstruction. The modified wavelet coefficients enter to the inverse DWT. Signal details contained in vectors $d^1, \ldots, d^{11}$ are zero-padded so the resulting length of signal in each frequency band is 2^{11}. Vectors with undesirable frequencies set to zeros are following:

$$r^{(11)*} = [0, \ldots, 0, d_1^1, \ldots, d_{1024}^1],$$
$$r^{(10)*} = [0, \ldots, 0, d_1^2, \ldots, d_{512}^2, 0, \ldots 0],$$
$$\ldots$$
$$r^{(1)*} = [0, d^{11}, 0, \ldots, 0],$$
$$r^{(0)*} = [a^{11}, 0, \ldots, 0].$$

Inverse discrete wavelet transform is applied to modified vectors:

$$s^{(i)} = \mathrm{iDWT}(r^{(i)*}).$$

MLP with three layers designed above is used for $s^{(i)}$ prediction. The patterns are given by the following way:

$$([s_1^{(i)}, s_2^{(i)}, \ldots, s_6^{(i)}], s_7^{(i)}),$$
$$([s_2^{(i)}, s_3^{(i)}, \ldots, s_7^{(i)}], s_8^{(i)}),$$
$$\ldots,$$
$$([s_{p-6}^{(i)}, s_{p-5}^{(i)}, \ldots, s_{p-1}^{(i)}], s_p^{(i)}),$$

where $i = 1, \ldots, n$ and p is the number of learning set patterns. Predicted values in given frequency bands are $\hat{s}_{p+1}^{(i)}, \hat{s}_{p+2}^{(i)}, \hat{s}_{p+3}^{(i)}$, where $i = 1, \ldots, n$. Then the m-steps forecast of original value x is given by the evaluation of

$$\sum_{i=1}^{n} \hat{s}_{p+m}^{(i)} + a^n$$

where a^n ($n = 11$ in our case) is the average in the last level of the wavelet decomposition.

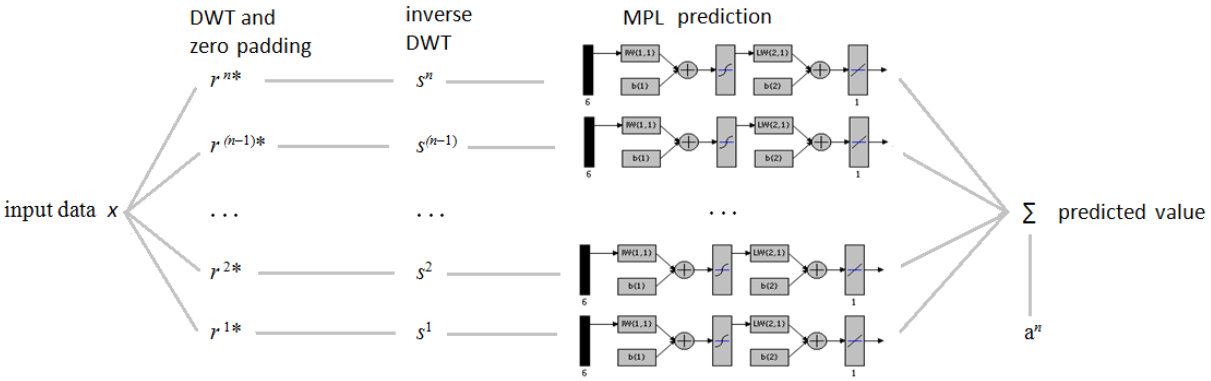


Figure 4 Structure of wavelet-neural system

As [1] shows the influence of low signal frequencies is minimal. The suggested algorithm contains the prediction of the four fastest frequency bands and estimation of additional values can be substituted by the previous values. The evaluation of vectors $r^{(8)*}, \ldots, r^{(11)*}$ is sufficient. Figure 5 illustrates this situation.

Figure 4 contains the time series processing scheme: iDWT of preprocessed data enters to the MLP designed according to Figure 1. Number of perceptrons in hidden layers for chosen decomposition levels (8, 9, 10, 11) are 6, 2, 1, and 1 respectively. Number of perceptrons in the first ANN layer is 6 for each level as Figure 1 illustrates. The last layer is single element usually.

The learning set for back-propagation algorithm contains the first 1800 patterns for each frequency range, and the testing set contains 100 patterns for each range from 8 to 11.

5 Conclusion

The Table 1 provides summary of exchange rate prediction errors (m = 1, 2, 3, 4). The predicted exchange rates were divided into four groups of the 1st, 2nd, 3rd and 4th forecasted rate. All these values were compared to the real exchange rates with the following error measures:

- RASE – root average square error,
- MAPE – mean absolute percentage error,
- DIR – the percentage the correct identification of the next value direction.

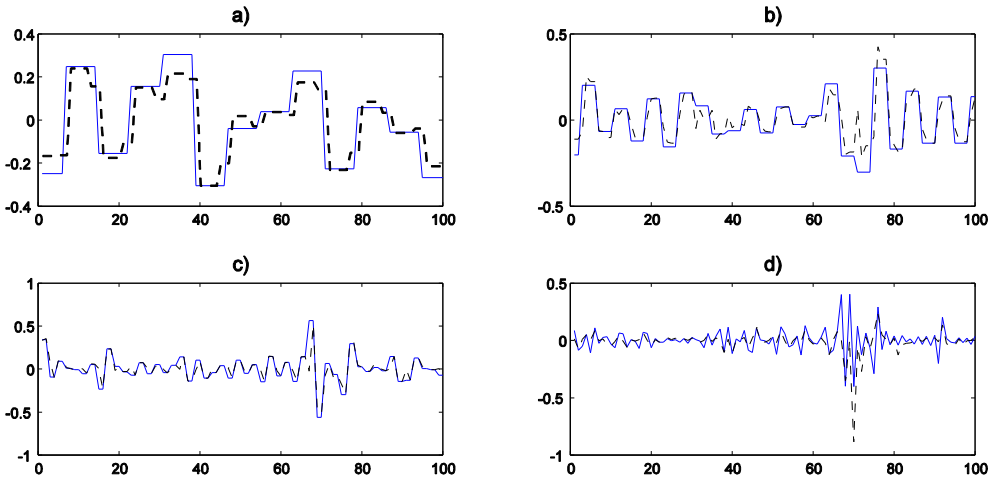


Figure 5 Signal prediction (Day 1) for the four highest frequency bands: a) k = 8, b) k = 9, c) k = 10, d) k = 11, (solid line – iDWT of zero padded wavelet coefficients, dashed line – its prediction by the MLP)

The graphs presented in Figure 5 indicate that the ANNs predict frequency bands pretty well. Surprisingly the RASE error for the whole signal decreases as the forecast period is extended to 3rd day. The MAPE, RASE errors are relatively high as well as direction coefficients from the 1st and 2nd forecasted exchange rate. DIR achieves the maximum value at the 3rd rate similar to [1].

Measure	Day 1	Day 2	Day 3	Day 4
RASE	0.438	0.433	0.424	0.587
MAPE	1.32 %	1.32 %	1.34 %	1.70 %
DIR	45 %	43 %	55 %	45 %

Table 1 Results of forecasting

The applied methods provide the results presented. The wavelet-neural system is the efficient tool for time series forecasting. Further optimization of ANNs training could improve the final prediction results.

Acknowledgements

The paper was supported by the grant F4/14/2011 from University of Economics in Prague.

References

[1] Bielecki, A., Hajto P., and Schaefer R.: Hybrid neural systems in exchange rate prediction. *Studies in Computational Intelligence* (*SCI*) **100** (2008), 211–230. Springer-Verlag Berlin Heidelberg, 2008.
[2] Burrus, C. A., Gopinath, R. A., and Guo, H.: *Wavelets and wavelet transform*. Prentice Hall, New Jersey, 1998.
[3] Haykin, S.: *Neural networks. A comprehensive foundation*. Pearson Education, Singapore, 2005.
[4] Mallat, S. G.: A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transaction on Pattern Analysis and Machine Intelligence* **11** (1998), 674–693.

Possibilities of computable general equilibrium techniques for analysis of the impact of selected government policies using the CGE of the Czech Republic

Václav Školuda[1]

Abstract. The presented paper deals with the simulation of the impact of some government policies in the macroeconomic scale. The analysis uses the computable general equilibrium (CGE) macroeconomic model for the Czech Republic. CGE model requires for its application in practice two basic pillars. The first is the data base from which it will build and second is a set of equations describing the events in the selected parts of economy. Data base is usually called Social Accounting Matrix (SAM). Choice of equations depends on the author of a particular model and its focus. In our case we will discuss the economy of whole country with regard to foreign trade. The first part of the paper consists of familiarizing the reader with the problems of general equilibrium models. The second part demonstrates the creation and calibration of the model. In the third part is the particular model used to achieve the objectives of the work. In this case an analysis of macroeconomic impacts of the changes in taxes and export-import policy. Several scenarios will be used to explain the possible consequences of actions. The conclusion will summarize the recommendations obtained by the quantitative analysis.

Keywords: general equilibrium, CGE model, SAM matrix

JEL Classification: C68, D58
AMS Classification: 91B5

1. Computable general equilibrium models

Models of equilibrium have recently become very popular, which is partly caused by the development of computing applications and simplifying the calculations. The presented article deals with the construction and application of a specific model of general equilibrium. Unlike partial equilibrium models that assume only equilibrium on partial markets, general equilibrium models assume equilibrium throughout the economy. Into this steady state is then introduced a shock and we monitor the impact on the economy. Impacts can be viewed as on the aggregate level (households, government) so at the level of different production sectors. It all depends on the level of aggregation in the model. In our case, we will focus on the automotive industry. CGE models are very popular in government institutions (for example [2] and [10]) because of revealing the unexpected effects of various interventions in the economy, or in other words, it reveals the impact of shocks on the economy as a whole and on its various parts. The submitted model was created completely independently from other empirical models used in the economy and is based on data specially created for this purpose. CGE model is usually described as an integrated system of simultaneous nonlinear equations derived from microeconomic theory to optimize the behavior of all economic agents (consumers, government ...), which tries to capture all transactions that take place between these agents and to bring the steady numerical solution. General equilibrium models are in many ways the opposite of econometric models (in particular VAR), since they are based on a strong theoretical foundation. Econometric estimates are sometimes used to estimate parameters of production functions. However, the sizes of these parameters are mostly taken from literature or studies focusing on this topic.

1.1 Social accounting matrix

For successful creation of general equilibrium model are data needed (as for each analysis). In the case of CGE models there is a need for specific configuration of data prior to the analysis in the so-called Social Accounting Matrix. This matrix shows balanced nominal flows in the economy of a country at some period, usually a year. In other words it is a complex structure involving all transactions in a given country and year. Social Accounting Matrix is used in modeling to calculate the initial equilibrium. The rows of such matrices represent income and columns correspond to expense (the principle of input-output). Of course, the totals for individual lines must be equal to the amount of the corresponding columns (the principle of national accounts). This is achieved not so easy as it seems and so is for the Czech Republic only available data for the years 2004 to 2006. For other coun-

[1] University of economics, Prague, Department of econometrics, W. Churchill sq. 4, Prague, vaclav@skoluda.sk.

tries it is easiest to mine data from the GTAP[2] project, which has the latest data from 2007. Because of data availability the year 2009 will be analyzed. For this year, the latest data are available on the pages of the Czech Statistical Office (ČSÚ), in particular symmetric input-output tables. Another additional source is a database of Eurostat and, in particular national accounts. For more on the issue look for example [1], [3] or [7].

1.2 Equations of the model

General equilibrium model describes the events as a whole. Therefore, it should also define the equations as a whole for this purpose. If the model describes the whole country in one year, we have to define all the relationships that occur in the national economy. From a mathematical point of view, we are using to achieve this, the individual equations. The equations are based on microeconomic theory and describing demand of firms, households, governments and investment, abroad, the balance of payments, zero earnings, income of households, investments and government and the balance of the markets.

Production functions

The model is using Leontieff production function, Cobb-Douglas production function, CES and CET production function. For all these functions it is necessary to derive the conditional demand, which will be used in the model specification. For completeness here are the functions followed by their conditional demands.

Leontieff production function (1) with elasticity of substitution σ=0 and its conditional demand (2):

$$Y = \gamma \min\left(\frac{X_1}{\alpha_1},...,\frac{X_n}{\alpha_n}\right) \quad where \quad \sum_{i=1}^{n}\alpha_i = 1 \tag{1}$$

$$\hat{X}_j = \frac{\alpha_j}{\gamma} Y \tag{2}$$

Cobb-Douglas production function (3) with elasticity of substitution σ=1 and its conditional demand (4):

$$Y = \gamma \prod_{i=1}^{n} X_i^{\alpha_i} \quad where \quad \sum_{i=1}^{n}\alpha_i = 1 \quad a \quad \gamma > 0 \tag{3}$$

$$\hat{X}_j = \frac{Y\alpha_j}{P_j\gamma}\prod_{i=1}^{n}\left(\frac{P_i}{\alpha_i}\right)^{\alpha_i} \tag{4}$$

CES production function (5) and its conditional demand (6):

$$Y = \gamma\left(\sum_{i=1}^{n}\alpha_i X_i^{\rho}\right)^{\frac{1}{\rho}} where \quad \sum_{i=1}^{n}\alpha_i = 1 \quad \gamma \geq 0 \quad \sigma = \frac{1}{1-\rho} \quad \rho \leq 1 \quad \rho \neq 0 \tag{5}$$

$$\hat{X}_j = \frac{Y}{\gamma}\left(\frac{P_j}{\alpha_j}\right)^{\frac{1}{\rho-1}}\left(\sum_{i=1}^{n}\alpha_i\left(\frac{P_i}{\alpha_i}\right)^{\frac{\rho}{\rho-1}}\right)^{-\frac{1}{\rho}} \tag{6}$$

CET production function has the same prescription as CES, differs only in the condition ρ≥1

Firms demand

Firms in the model are aggregated into four sectors, labeled *ag* – agriculture, *in* – industry, *au* – automotive, *sr* – services. Indexed *i* or *j* for rows or columns. Each of these sectors uses inputs *L* – labor, *K* – capital and X_j – commodities of sectors. Model assumes behavior of firms according to Cobb-Douglas production function, the conditional demand for labor (7), capital (8) and j-th commodity (9) can be written as follows, where w^E is the price of labor in the sector, *r* is the cost of capital and P^{DS} vector of commodity prices on the domestic market:

$$L_i = L_i\left(Y_i, w^E, r, P^{DS}\right) \tag{7}$$

[2] Global Trade Analysis Project (GTAP) is a worldwide group focused on quantitative analysis of global trade. It has data on trade throughout the world and on this basis also performs analysis by CGE models.

$$K_i = K_i\left(Y_i, w^E, r, P^{DS}\right) \tag{8}$$

$$X^i_j = X^i_j\left(Y_i, w^E, r, P^{DS}\right) \tag{9}$$

Demand of households, government and investment

The model is using Cobb-Douglas function, and if the households behave rational, their demand can be derived with additional production sectors making up the aggregate commodity TH - household wealth. Similarly, demand is solved by TG – government, TINV - investment.

$$H_j = H_j\left(TH, P^{DS}\right) \tag{10}$$

$$G_j = G_j\left(TG, P^{DS}\right) \tag{11}$$

$$INV_j = INV_j\left(TINV, P^{DS}\right) \tag{12}$$

Abroad

Model assumes a different nature of the products from domestic production and other from abroad so the rate of substitution is low. The total supply of domestic goods is thus modeled by CES function of domestic and foreign commodities. Domestic producers are deciding between domestic and foreign markets and we model them by CET. Where IM_j – the amount of imported commodity j, EX_i – exported commodity, DS_j – domestic supply of commodity, DP_i – domestic production for the domestic market, P^{IM} –price of imports, P^{EX} –price of exports.

$$DP_j = DP_j\left(DS_j, P_j, P^{IM}_j\right) \tag{13}$$

$$IM_j = IM_j\left(DS_j, P_j, P^{IM}_j\right) \tag{14}$$

$$DP_i = DP_i\left(Y_i, P_i, P^{EX}_i\right) \tag{15}$$

$$EX_i = EX_i\left(Y_i, P_i, P^{EX}_i\right) \tag{16}$$

Balance of payments

This equation expresses the deficit or surplus from account balance and the model assumes that the balance of payments bear the investment. Where ER – exchange rate and t^X_Y – transfers between sectors; from sector X to Y:

$$BP = \sum_i P^{IM}_i IM_i - \sum_j P^{EX}_j IM_j - ER\left(t^L_{ROW} - t^{ROW}_L + t^H_{ROW} - t^{ROW}_H + t^G_{ROW} - t^{ROW}_G\right) \tag{17}$$

Equations for zero profit and budget limitations

Production sectors meet the conditions of zero profit, i.e. net revenues must equal costs. These conditions meet also households, government and investment. There should be also fulfilled and budget limitations. Where P^{TH} – price level of total household consumption (welfare), P^{TG} – government, P^{TINV} – investment.

$$P_i DP_i + P^{EX}_i EX_i = w^E L_i + rK_i + \sum_j P^{DS}_j X^i_j \tag{18}$$

$$P^{DS}_j DS_j = P^{IM}_j IM_j + P_j DP_j \tag{19}$$

$$P^{TH} TH = \sum_j P^{DS}_j H_j \quad ohr. \quad P^{TH} TH = b^H_C M^H \tag{20}$$

$$P^{TG} TG = \sum_j P^{DS}_j G_j \quad ohr. \quad P^{TG} TG = b^G_C M^G \tag{21}$$

$$P^{TINV} TINV = \sum_J P^{DS}_j INV_j \quad ohr. \quad P^{TINV} TINV = M^{INV} \tag{22}$$

Income of households, government and investment

Household income (23) consists of wages, capital and transfers. Government revenues (24) consist of tax, capital and transfers. Income of investment (25) is formed by transfers and resources, which households and government misses for consumption. Where b_K^H – capital owned by households, b_K^G – capital owned by government, b_C^H – marginal propensity to consume of households, b_C^G – marginal propensity to consume of government.

$$M^H = \sum_i wL_i + b_K^H \sum_i rK_i + t_H^G + t_G^H + ER\left(t_L^{ROW} - t_{ROW}^L + t_H^{ROW} - t_{ROW}^H\right) \tag{23}$$

$$M^G = \sum_i t_i wL_i + b_K^G \sum_i rK_i - t_H^G + t_G^H + ER\left(t_G^{ROW} - t_{ROW}^G\right) \tag{24}$$

$$M^{INV} = \left(1 - b_C^H\right)M^H + \left(1 - b_C^G\right)M^G + ER\left(t_{INV}^{ROW} - t_{ROW}^{INV}\right) \tag{25}$$

Balance of the markets

If in the economy does not exist unsaturated demand (Walras law), shall equal the supply and demand in all markets. In addition to the balance of payments equilibrium (17) also provides for equality in the commodity market (26). Other balances are related to the closure model.

$$DS_j = \sum_j X_j^i + H_j + G_j + INV_j \tag{26}$$

$$TK = \sum_i K_i \tag{27}$$

1.3 Choice of the model closure

Model as defined so far contains more variables than equations. It is therefore necessary to define model and close by fixing the size of two variables from a set of TK - total capital, TL - total labor and TINV - total investment. And this choice is on the creator of the model [4]. Classical macroeconomic theory proposes fixing of total supply of capital and assumes full employment. Keynesian closure assumes unemployment and therefore in addition to the overall supply of capital sets a fixed level of total investment. Another option, in practical models almost unused, is the use of so-called alternative closure [9]. This closure defines the auxiliary variable, which tries to minimize deviations in the weights set. Basically it is looking for something between classical and Keynesian closure. From these it takes its common sign in the form of fixing the overall supply of capital (27) and solves the following minimization task, where *bTINV* – initial value of total investments from the calibration of model, *bTL* – initial value of labor supply from the calibration.

$$e = \left(\frac{TINV}{bTINV} - 1\right)^2 + \left(\frac{TL}{bTL} - 1\right)^2 \rightarrow \min \tag{28}$$

by constrains with all model equations.

1.4 Calibration of the model

Running the model and obtaining the first results prior calibration of the model itself. The calibration consists in quantifying the parameters of model equations. Individual parameters can be calibrated based on data contained in the SAM, based on the literature, using expert estimate or using econometric modeling. Most of the values are obtained from the Social Accounting Matrix. More for example [8] and [11], on which is also based on this model.

2 General equilibrium model of the Czech Republic

The aim of this paper is to evaluate the effects of selected shocks on the entire economy of the Czech Republic. For this purpose, it is created a general equilibrium model that takes into account some specific features of the Czech Republic, like the choice of production functions and elasticities, as well as car production sector allocation of aggregate industry sector. General equilibrium models have the indisputable advantage in that they are describing economy as a whole and thus it is possible to track the impact of any action on its part. The disadvantage of these models is that they provide only a relative (percentage) change of variables, which is principally due to the fact that we compare two equilibrium states, before and after shock in the economy. Specific scenarios of the impact of changes in the taxes and exports restrictions are part of the results.

2.1 Social accounting matrix of Czech Republic

Entering data for a general equilibrium model are data in specific form of the so-called Social Accounting Matrix (SAM). Table 1 shows the SAM for the Czech Republic for 2009 in millions of Czech crowns. Due to the absence of a suitable matrix, this was created specifically for this model. Aggregation of data and balancing across the table are the work of the author. The data in the matrix are nominal flows in the economy. The shortcuts in table accords: *ag* – agriculture, *in* – industry, *au* – automotive, *sr* – services, *L* – labor, *T* – taxes, *K* – capital, *H* – households, *G* – government, *INV* – investment, *ROW* – Rest of the world. The lines represent income and columns expenditure. The cell SAM[au, ROW] shows for example the transfers which the automotive industry is receiving from abroad, i.e. export. Values on the diagonal of SAM represent intermediate consumption, for example the cell SAM[ag, ag] corresponds to intermediate consumption in agriculture.

	ag	in	au	sr	L	T	K	H	G	INV	ROW
ag	23 436	299 256	979	23 896				38 529	485	3 180	48 057
in	76 042	1 908 399	214 128	680 728				481 931	32 616	551 952	1 367 744
au	1 378	29 299	191 681	50 601				41 473	858	71 351	411 439
sr	40 514	518 061	72 135	1 498 732				819 019	674 929	135 766	341 295
L	50 053	559 303	60 360	1 075 335							16 719
T	4 103	31 719	1 647	275 397							
K	72 154	524 615	51 823	195 554							
H					1 723 425		582 812		186 162		154 340
G						312 866	261 334	336 802			60 873
INV								902 879	40 415		83 972
ROW	170 137	1 442 888	205 327	300 208	38 346			26 105	36 410	265 017	

Table 1 Social accounting matrix of Czech Republic, 2009. *Source: ČSÚ, Eurostat, own calculations*

2.2 Construction of the model

The actual model is by its very nature a nonlinear optimization problem. To solve this task are used computers, in this case the GAMS environment. This program allows writing equations in algebraic form and solving them with some of the built-in solvers. Presented model is constructed by the author and named VseMoR - General Equilibrium Model constructed at the University of Economics in Prague. The construction of the model consists of loading data from SAM, defining of indices, parameters and variables, the model equations assembly, and finally the solution of the model. The size of parameters is loaded either directly from the Social Accounting Matrix, or by other means described in the section on model calibration. Putting the list of endogenous variables is essential for the solution of the model. Notation of equations consists of compilation of equations, as shown in the model equations section. The solution lies in putting starting values for the solver and the selection of a specific solver. Construction of model is followed by the definition of the selected scenarios of introducing shocks to the economy. After these shocks the economy reaches a new equilibrium, resulting in a comparison of changes in value of the endogenous variables in the model, and these changes are expressed as a percentage. Technical aspect of the model shows VseMoR as a record of GAMS commands of approximately 500 lines. The actual model consists of 89 equations. The model is static and open; the code itself is based on the works [5], [9] and [11]. By the construction of the model it was kept in mind on its ease of scalability and it is thus easy to adjust for other specific analysis or for use with another database, whether differing in the degree of aggregation or based on another country. Also, it is not technically difficult to adjust to a dynamic model (recursively). The author considers the results of such a model to be inadequate due to the fact that on the basis of these models is commonly performed the long term forecast (up to 20 years), but the data base itself remains unchanged and the model essentially assumes equilibrium in a single year.

3 Results

The aim was next to building the model and verifying its functionality also the quantification of the relative impact of selected government policies. Observed are the impacts of changes in tax rates and the import-export regulations of the government. Impacts of selected shocks are observed both individually and simultaneously. A well-constructed model allows essentially monitor any shocks in the economy and its impacts on any part thereof. Of course if this part is included in the model.

Scenario1	Scenario2	Scenario3	Scenario4	Scenario5	
-3.08	2.54	-0.09	4.70	4.32	% change in total consumption of Households
-3.91	-0.57	-2.50	0.83	0.43	% change in total consumption of Government
1.70	5.46	2.04	8.11	12.27	% change in total consumption of Investment
-1.26	2.63	0.29	4.93	7.27	% change in total Import
-1.23	2.57	0.28	4.82	7.10	% change in total Export
-1.13	2.36	0.21	4.13	4.02	% change in production in sector agriculture
-0.99	3.07	0.32	4.76	5.42	% change in production in sector industry

-0.50	0.02	0.80	5.08	10.95	% change in production in sector automotive
-3.08	1.71	-0.55	4.68	3.52	% change in production in sector services

Table 2 Results of various scenarios

3.1 Scenarios

Construction of the model is followed by application of selected scenarios. Summary of the results provides Table 2. All data are percentages and expressed as percent change of selected variables. The **first** reference scenario is flat tax increase of 10%. The table shows the results column s1. After the shock for example consumption of investment has increased by 1.7% and output in service sector fell by 3%. The **second** scenario assumes a total reduction of exports in the economy by 5%. This reduction has an impact in total government consumption by reduction of 0.5% and for example on automotive industry has almost no impact. The **third** scenario is about to reduce taxes in the automotive industry (producers of cars) by 20%. The purpose of this scenario is to monitor the impact of the so-called car scrapping in the economy. The results show that this change is least affected by the economy and major changes can be observed only in reducing the overall government consumption and increase in consumption of investment. The **fourth** scenario is fictitious increase in compensation of employees across the country by 3%. The result of this change is of course a big increase in consumption and also production in all sectors ranging from 4 to 5%. The **fifth** and last presented scenario is combination of the first two scenarios. It is therefore an increase in taxes by 10% and a simultaneous decrease of exports by 5%. The impact of these interventions is present in two areas by more than two digits. Total consumption by investment will increase in this case by over 12% and rise in car production is nearly 11%. These scenarios and their impacts have of course just an illustrative nature. The model by its very nature can show impacts of any scenario on any part of the economy, which is included in it. All scenarios apply for other conditions unchanged (ceteris paribus).

4 Conclusion

General equilibrium models are part of macroeconomic models. They are based on strong microeconomic foundations which consist of system of equations attempting to describe the whole economy and achieve so to macroeconomic model. This paper presents construction of such smaller model and shows the possibilities of its application. The finished model was applied to the Czech Republic. The data source was for this purpose constructed Social Accounting Matrix. Examples of impacts of various scenarios affecting the economy are summarized in Table 2. Model VseMoR is unlike other similar models and analyzes based on them relatively independent. It is based on its own processed data from different sources. Equation of the model are designed and calibrated for specifics of the Czech Republic. The model does not expect the behavior of classical or Keynesian economics, but works somewhere in between. By developing the model it was kept in mind also its future use, so it is easily customizable to changing input conditions. All these features make this model unique.

Acknowledgements

This paper was created within the project of Internal Grant Agency IGA F4/1/2012.

References

[1] Beutel, J. et al.: *Eurostat Manual of Supply, Use and Input-Output Tables*. European Commission, Luxembourg, 2008.
[2] Bora, B., Cernat, L., Turrini, A.: *Duty and Quota Free Market Access for LDCs: Further Evidence from CGE Modeling*. United Nations, Geneva, 2002.
[3] Cvengroš, F.: *Ekonomické tabulky (SAM) a modely pro výpočet všeobecné rovnováhy (CGE)*. ÚÚNV, Praha, 1991.
[4] Decaluwé, B., Martens, A., Monette, M.: *Macroclosures in Open Economy CGE Models: A Numerical Reappraisal*. Université de Montréal, Montreal, 1987.
[5] Drozdík, J.: *Všeobecný algebraický modelovací systém – GAMS.* Diplomová práca, FMFI UK, Bratislava, 2006.
[6] Herr, B., Maussner, A.: *Dynamic General Equilibrium Modeling.* Springer, Berlin, 2009.
[7] Hronová, S. et al.: *Národní účtníctví: Nástroj popisu globální ekonomiky*. C. H. Beck, Praha, 2009.
[8] Křístková, Z.: *Simulace dopadu nástrojů zemnědělské politiky na českou ekonomiku v kontextu obecné rovnováhy.* Dizertačná práca, PEF CZU, Praha, 2009.
[9] Pániková, L.: *Alternatívne uzávery CGE modelov*. Diplomová práca, FMFI UK, Bratislava, 2007.
[10] Vaittinen, R.: *Trade policies and integration – evaluations with CGE-models.* Helsinki School of Economics, Helsikni, 2004.
[11] Vojteková, M.: *Analýza zmien dôchodkového system pomocou CGE modelov*. Diplomová práca, FMFI UK, Bratislava, 2009.

Tobin tax introduction and risk analysis in the Java simulation

Roman Šperka[1], Marek Spišák[2]

Abstract. This paper deals with the agent-based simulation. The aim of the paper supported by the hypothesis is to prove positive impact of Tobin tax introduction together with the risk analysis on the stability of financial market. The core of the paper is the Java implementation of the multi-agent system being served as simulation framework. Multi-agent systems consist of a significant number of cooperating elements, so called intelligent agents. A multi-agent financial market model and simulation is introduced. Intelligent agents follow technical and fundamental trading rules to determine their speculative investment positions. We consider direct interactions between speculators due to which they may decide to change their trading behaviour. For instance, if a technical trader meets a fundamental trader and they realize that fundamental trading has been more profitable than technical trading in the recent past, the probability that the technical trader switches to fundamental trading rules is relatively high. In particular the influence of transaction costs and risk is studied.

Keywords: ABMS, simulation, Tobin tax, risk analysis, financial model.

JEL Classification: E37, E44, C53, C63, C90
AMS Classification: 68U20

1 Introduction

Intelligent agent technology used in this paper has deeper roots in economic theory history, mainly in the ideas of F.A. Hayek and H.A. Simon.

One of the main ideas of F.A. Hayek is that the economic system should be studied from bottom. He stresses the need to look at the market economy as to a decentralized system consisting of mutually influencing individuals (the same goes for financial markets) in his work. In "Individualism and Economic Order" Hayek [4] writes: "There is no other way to understand social phenomena such as through our understanding of the actions of individuals who are oriented towards other people and management according to their expected behaviour." He opposed mainly against collectivist theories which claim to be able to fully understand the social right, regardless of the individuals who constitute them. Hayek also elaborated on the theme of dispersed interactions between individuals. The basis for thinking about human society is the fact that no member can know more than a tiny part of society and therefore everything that goes into the decision-making are the immediate results that will have proceedings to the surroundings. This approach builds a contrast with the assumption of perfect information, which is used in traditional equilibrium analysis. In the theory of complex systems, where Agent-based Modelling and Simulation (ABMS) clearly falls, is this idea the primary principle [8]. Agents, unlike classical equilibrium approach have not perfect information about all processes in the system.

No strict rules are conducted to the agents. They themselves select those practices that lead to the best results according to the success of strategies and rules. They are not looking for universal general rule. They are governed by a method that has proven in the environment under given conditions in the past. Multi-agent approach uses various scientific methods for introducing the adaptive behaviour of the program structures (e.g. [5]). The basic feature of complex adaptive systems is that their global properties can be easily derived from the characteristics of individual units (agents). Although each agent structure is simple, the behaviour of the system as a whole can be very difficult. A complex system is not the same as a chaotic system. Generally, a complex system tends to evolve away from both extremes - full of randomness on the one hand and absolute order on the other [6]. Multi-agent systems are based on the selection of behaviour rules that are subjectively optimal in certain environment for each agent functioning. Multi-agent system implemented through ABMS consists of two types of rules - spontaneous and created. The agent should be determined what their purpose is, what variable or group of variables to be monitored and optimized. On the other hand the way for reaching goals, is already left full of them.

[1] Silesian University in Opava, School of Business Administration in Karviná, sperka@opf.slu.cz
[2] Silesian University in Opava, School of Business Administration in Karviná, spisak@opf.slu.cz

Computational social science involves the use of ABMS to study complex social systems [3]. ABMS consists of a set of agents and a framework for simulating their decisions and interactions. ABMS is related to a variety of other simulation techniques, including discrete event simulation and distributed artificial intelligence or multi-agent systems [7]. Although many traits are shared, ABMS is differentiated from these approaches by its focus on finding the set of basic decision rules and behavioural interactions that can produce the complex results experienced in the real world [9]. ABMS tools are designed to simulate the interactions of large numbers of individuals so as to study the macro-scale consequences of these interactions [11]. Each entity in the system under investigation is represented by an agent in the model. An agent is thus a software representation of a decision-making unit. Agents are self-directed objects with specific traits and typically exhibit bounded rationality, that is, they make decisions by using limited internal decision rules that depend only on imperfect local information. In practice, each agent has only partial knowledge of other agents and each agent makes its own decisions based on the partial knowledge about other agents in the system.

The transaction costs on the financial market are mainly the costs of the obtaining and the interpreting of the information, the time required for decision making, various types of fees, etc. Transaction costs according to Burian [2] are often viewed as negative phenomena, but there are cases where the increase in the transaction costs can be viewed positively and can contribute to the stability of the market. The increase in the transaction costs may also occur in the form of non-market regulation such as the taxes. In the early seventies the Nobel laureate in the economics James Tobin drafted the regulation of currency markets. Tobin suggested that all short-term transactions should be taxed at a low fixed rate (the proposal was later identified as the so-called Tobin tax). The results according to Tobin would avoid short-term currency speculation and stabilize the market. Currency speculation can lead to the sudden withdrawal of the currency from the circulation in order to artificially increase the price. The consequence for the economy of the countries that use this currency may be a temporary reduction in liquidity, problems in obtaining loans and other phenomena that can lead to the reduced growth or even to the recession. Tobin tax was never implemented.

For our research work, a multi-agent system will be implemented which is able to deal with unpredictable phenomena surrounding every company nowadays able of agents behaviour investigation. Multi-agent system will be developed and managed as a simulation framework in JADE development platform (JAVA programming language [1]). This paper deals with the agent-based simulation of the Tobin tax introduction together with the risk analysis and their impact on the stability of financial market. The motivation is to investigate the reaction of financial market on the higher transaction costs and risk application. A multi-agent financial market model and simulation is introduced. Intelligent agents follow technical and fundamental trading rules to determine their speculative investment positions. We consider direct interactions between speculators due to which they may decide to change their trading behaviour. For instance, if a technical trader meets a fundamental trader and they realize that fundamental trading has been more profitable than technical trading in the recent past, the probability that the technical trader switches to fundamental trading rules is relatively high. In particular the influence of transaction costs and risk is studied. The aim of the paper supported by the hypothesis is to prove the positive influence of Tobin tax introduction on the stability of financial market. This paper is structured as follows. Section 2 firstly describes the original mathematical model, secondly informs about previous simulation results, and lastly represents the hypothesis. Section 3 presents the original simulation results of the agent-based model of financial market.

2 Model description

2.1 Original model description

The model developed by Frank Westerhoff [12] was chosen for the implementation. It is an agent-based model, which simulates the financial market. Two base types of traders are represented by agents:

- **Fundamental traders -** their reactions are based on fundamental analysis – they believe that asset prices in long term approximate their fundamental price – they buy assets when the price is under fundamental value.
- **Technical traders -** decide using technical analysis – prices tend to move in trends – by their extrapolating there comes the positive feedback, which can cause the instability.

Price changes are reflecting current demand excess. This excess is expressing the orders amount submitted by technical and fundamental traders each turn and the rate between their orders evolves in a time. Agents regularly meet and discuss their trading performance. One agent can be persuaded to change his trading method, if his rules relative success is less than the others one. Communication is direct talk one agent with other. Communicating agents meet randomly – there is no special relationship between them. The success of rules is represented by current and passed profitability. Model assumes traders ability to define the fundamental value of assets and the agents behave rationally.

The price is reflecting the relation between assets that have been bought and sold in a turn and the price change caused by these orders. This can be formalized as a simple log-linear price impact function.

$$P_{t+1} = P_t + a(W_t^C D_t^C + W_t^F D_t^F) + \alpha_t \quad (1)$$

Where a is positive price adjustment coefficient, D^C are orders generated by technical angents while D^F are orders of fundamental ones. W^C and W^F are weights of the agents using technical respectively fundamental rules. They are reflecting current ratio between the technical and fundamental agents. α brings the random term to the Figure 1. It is an IID[3] normal random variable with mean zero and constant standard deviation σ^α.

As was already said, technical analysis extrapolates price trends – when they go up (price is growing) agents buy the assets. So the formalization for technical order rules can be like this:

$$D_t^C = b(P_t - P_{t-1}) + \beta_t \quad (2)$$

The parameter b is positive and presents agent sensitivity to price changes. The difference in brackets reflects the trend and β is the random term – IID normal random variable with mean zero and constant standard deviation σ^β.

Fundamental analysis permits the difference between price and fundamental value for short time only. In long run there is an approximation of them. So if the price is below the fundamental value – the assets are bought and vice versa – orders according fundamentalists are formalized:

$$D_t^F = c(F - P_t) + \gamma_t \quad (3)$$

The parameter c is positive and presents agent sensitivity to reaction. F represents fundamental value – we keep as constant value to keep the implementation as simple as possible[4] . γ is the random term – IID normal random variable with mean zero and constant standard deviation σ^γ.

If we say that N is the total number of agents and K is the number of technical traders, then we define the weight of technical traders:

$$W_t^C = K_t/N \quad (4)$$

And the weight of fundamental traders:

$$W_t^F = (N - K_t)/N \quad (5)$$

Two traders meet at each step and they discuss about the success of their rules. If the second agent rules are more successful, the first one changes its behaviour with a probability K. Probability of transition is defined as $(1-\delta)$. Also there is a small probability ε that agent changes his mind independently. Transition probability is formalized as:

$$K_t \begin{cases} K_{t-1}+1 \ \ with\ probability \ \ p_{t-1}^+ = \dfrac{N-K_{t-1}}{N}\left(\varepsilon + (1-\delta)_{t-1}^{F\to C}\ \dfrac{K_{t-1}}{N-1}\right) \\ K_{t-1}-1 \ \ with\ probability \ \ p_{t-1}^- = \dfrac{K_{t-1}}{N}\left(\varepsilon + (1-\delta)_{t-1}^{C\to F}\ \dfrac{N-K_{t-1}}{N-1}\right) \\ K_{t-1} \qquad with\ probability \ \ 1 + p_{t-1}^+ - p_{t-1}^- \end{cases} \quad (6)$$

Where the probability that fundamental agent becomes technical one is:

$$(1-\delta)_{t-1}^{F\to C} = \begin{cases} 0.5 + \lambda\ for\ A_t^C > A_t^F \\ 0.5 - \lambda\ otherwise \end{cases} \quad (7)$$

Respectively that technical agent becomes fundamental one is:

$$(1-\delta)_{t-1}^{C\to F} = \begin{cases} 0.5 - \lambda\ for\ A_t^C > A_t^F \\ 0.5 + \lambda\ otherwise \end{cases} \quad (8)$$

Success (fitness of the rule) is represented by past profitability of the rules that are formalized as:

$$A_t^C = (\exp[P_t] - \exp[P_{t-1}])D_{t-2}^C + dA_{t-1}^C \quad (9)$$

[3] independent and identically distributed
[4] in our implementation F = 0

for the technical rules. And:

$$A_t^F = (\exp[P_t] - \exp[P_{t-1}])D_{t-2}^F + dA_{t-1}^F \quad (10)$$

for the fundamental rules. Agents use most recent performance (at the end of A^C formula resp. A^F) and also the orders submitted in period $t-2$ are executed at prices started in period $t-1$. In this way the profits are calculated. Agents have memory, which is represented by the parameter d. Values are $0 \leq d \leq 1$. If $d = 0$ then agent has no memory, much higher value is, much higher influence the profits have on the rule fitness.

2.2 Previous simulations

From our previous simulations [10] was seen that original model which was implemented has (in our parameterization) tendency to stabilize itself in a long term – if the fundamental trading rules are overbearing the technical trading method, although the bubbles and the crashes occur, their values are going to be smaller because the price is targeting near the fundamental value and the volatility is going to be less too.

After introduction of the transaction cost influence on the price – the price is going up to the bubble while technical traders are overtaking the market. Then possible two scenarios can occur:

- **Transaction costs value is low** – the price starts to be falling according the fundamental traders' weight growth. In this moment volatility falls down and the market stabilizes.
- **Transaction costs value is high** – fundamental traders' weight = 0, the system destabilizes and the price grows without limit.

2.3 Extension of original model and hypothesis

The risk was implemented as a price risk percentage (RP) which is generated each turn from given interval according uniform random distribution[5]. So for risk influence the price formula has changed in this way:

$$P_{t+1} = (P_t + a(W_t^C D_t^C + W_t^F D_t^F) + \alpha_t) * RP \quad (11)$$

Transaction costs were implemented in the same way as in previous simulations with adding constant value to the price:

$$P_{t+1} = (P_t + a(W_t^C D_t^C + W_t^F D_t^F) + \alpha_t) * RP + TC \quad (12)$$

The hypothesis is that transaction costs (12) will bring the same effect to the market as in the case of pure model without risk involvement – with small amount of TC it will stabilize the market [10]. Two types of simulations were done using (11) and (12) – one only with risk percentage and the second one with transaction costs to see the difference.

3 Simulation results

Simulation was done with 1 marketing agent and 500 trading agents. The rest of the parameters remained same as in original Westernhoff model:

$$a = 1,\ b = 0.05,\ c = 0.02,\ d = 0.95,\ \varepsilon = 0.1,\ \lambda = 0.45,\ \sigma\alpha = 0.0025,\ \sigma\beta = 0.025,\ and\ \sigma\gamma = 0.0025$$

With these parameters the model is calibrated to the daily data. Number of turns, resp. time steps is 360 days, which represents one year. Each generation (risk only and risk with transaction costs) was done 25 times. Results were aggregated to obtain more accurate results. Average data were calculated.

Interval for percentage values was decided as <-1, 3>; when 1 (as 100 %) means not changed price. Results can be seen in the following graphs. In Figure 1 and 2 on the top left position the price values can be seen. Top right graph represents changes of the price in a time. The bottom left graph shows the weights of technical trading rules (in a long time there is a tendency to prefer fundamental than technical trading rules). Bottom right graph includes the distribution of returns (log price changes) compared with the normal distribution. Figure 1 shows the results in case of risk involvement into model. All four parts of the figure describe current situation on the market. Volatility of prices is clearly visible. There is possibility that prices can change rapidly in both directions. In 2008 felt the prices significantly down and the market was destabilized. In order to prevent to the destabilization situation, we suggest Tobin tax implementation into the system.

[5] None single outcome has higher probability than others.

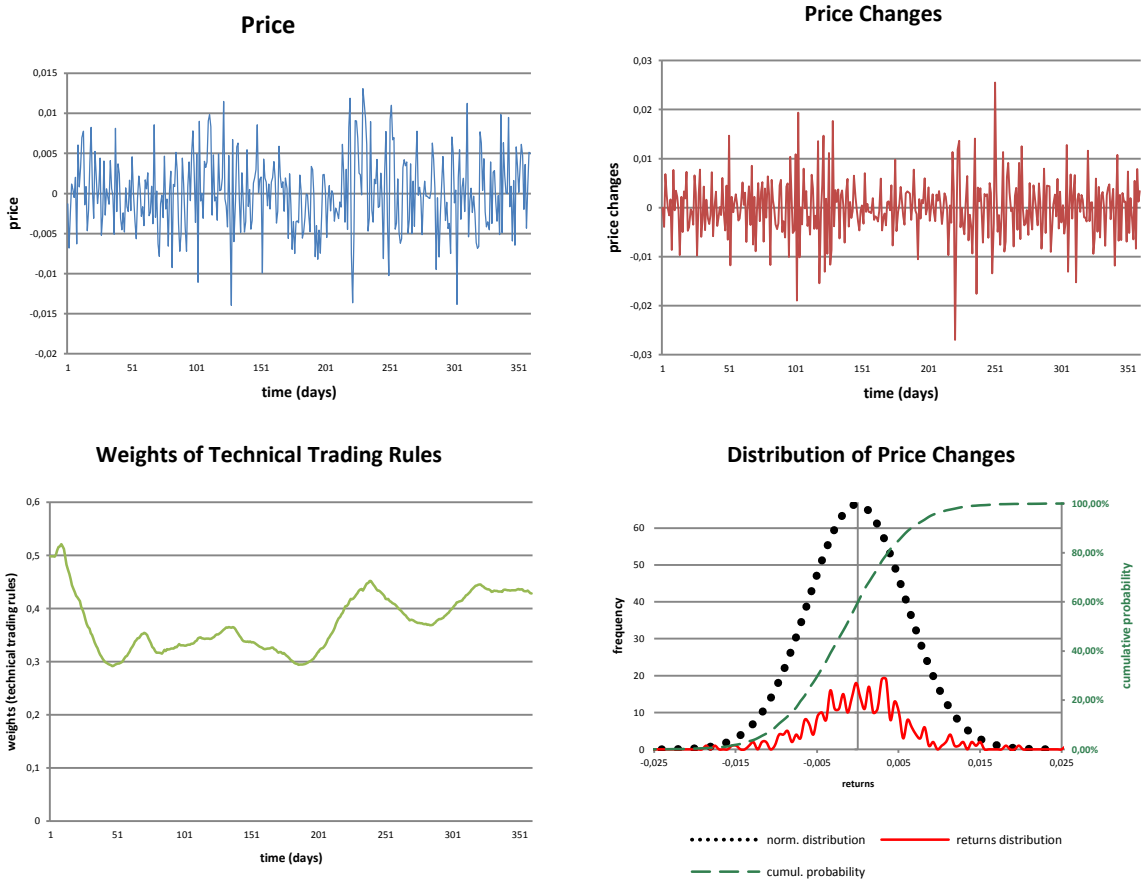
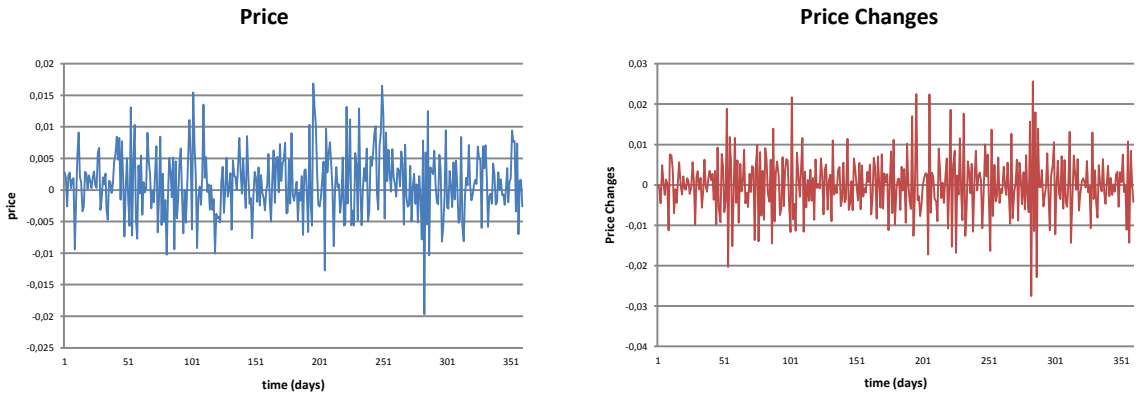


Figure 1 Results of risk involvement simulation (source: own)

We outcome from this situation and the Tobin tax implementation into the system was done. Tobin tax was involved in the model in the form of transaction costs. The value of the transaction costs was set to TC = 0.001. Results can be seen in the figure 2. The parts of the figure represent the same phenomena as in the above paragraph. Three parts of the figure seem to be same as in the risk involvement graph. Only one part has changed. In the long term the fundamental rules overcome the technical rules. That is one sign of the financial market stabilization. To confirm the hypothesis the price changes should be lower. This was not proven. So the hypothesis can be fulfilled only partially.



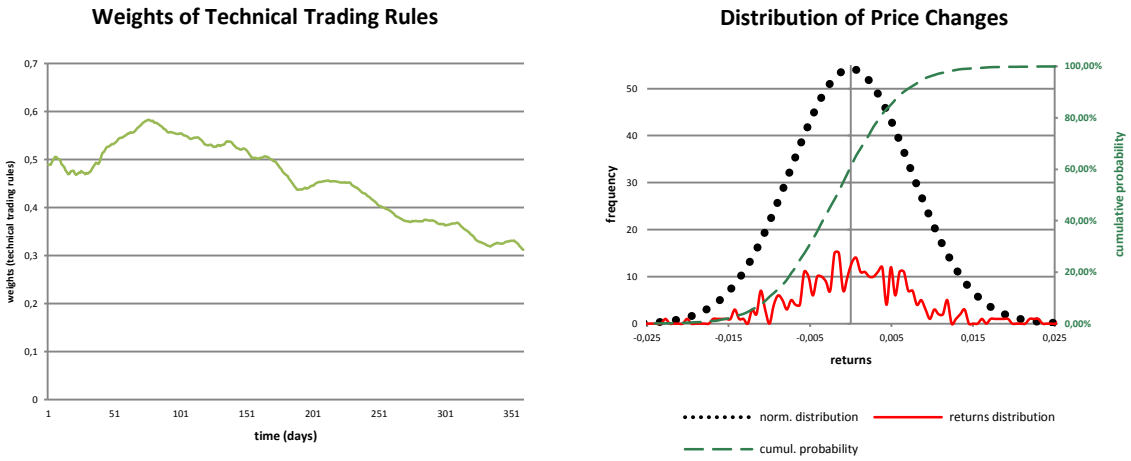


Figure 2 Results of risk involvement simulation with TC (source: own)

4 Conclusion

Agent-based simulation of financial market was introduced in this paper. Intelligent agents representing financial market participants followed fundamental and technical rules. The probability that agent switches from the fundamental to the technical behaviour depends on the historic trend of asset's prices. The hypothesis for our research was based on our previous simulation results proving that transaction costs influence (Tobin tax) stabilizes the financial market. We involved the risk into original model and we supposed that transaction costs introduction would lead to the predominance of fundamental rules, which will automatically cause price lowering and market stability (measured by volatility in price changes).

The hypothesis was fulfilled only partially – the fundamental rules have growing tendency in time, but the prices and their differences are nearly the same in both simulations.

We will focus on the risk and parametrization of the model in our next research steps in order to prove that Tobin tax has positive impact on the stability of financial market.

References

[1] Bellifemine, F., Caire, G., and Trucco, T.: *Jade Programmer's Guide. Java Agent Development Framework*, 2010. Available from WWW:<http://jade.tilab.com/ doc/programmersguide.pdf>.

[2] Burian, J.: *Multiagentní model transakčních nákladů na finančních trzích*. VŠE Praha, 2010. Available from WWW: < http://www.eldar.cz/honza/articles/burian_ace_fin_market_tc.doc >

[3] Epstein, J. M., and Axtell, R.: *Growing Artificial Societies: Social Science from the Bottom Up.* Brookings Institution Press, MA, 1996, ISBN: 0-262-55025-3.

[4] Hayek, F.A.: *Individualism and Economic Order*. Routledge and Kegan Paul Ltd., London, 1949.

[5] Holland, J.H.: *Adaptation in Natural and Artificial Systems*. The University of Michigan, 1975.

[6] Kochugovindan, S., and Vriend, N.J.: *Is the Study of Complex Adaptive Systems Going to Solve the Mystery of Adam Smith's 'Invisible Hand'?* The Independent Review, 3, No.1, 1998, 53-66.

[7] Law, A. M., and Kelton, W.D.: *Simulation Modeling and Analysis*. 3rd ed., McGraw-Hill: New York, 2000, ISBN 0070592926.

[8] Macal, C. M., and North, M. J.: Tutorial on Agent-Based Modeling and Simulation Part 2: How to Model With Agents. *Proceedings of the Winter Simulation Conference*, 2006, 73-83..

[9] Sallach, D. L., and Macal, C. M.: *Introduction: The Simulation of Social Agents*. Social Science Computer Review, Sage Publications, Thousand Oaks, CA, Fall 2001, vol. 19, no. 3, DOI 10.1177/089443930101900301, 15-29.

[10] Spišák, M., and Šperka, R.: *Financial Market Simulation Based on Intelligent Agents – Case Study*. Journal of Applied Economic Sciences, Volume VI, Issue 3(17), 2011, Romania, Print-ISSN 1843-6110.

[11] Tesfatsion, L.: *Agent-Based Computational Economics: Growing Economies from the Bottom Up*. Artificial Life, Vol. 8, No.1, MIT Press, Cambridge, MA, 2002, 55-82.

[12] Westerhoff, F.: *A simple agent-based financial market model: direct interactions and comparisons of trading profits. Working Paper No. 61.* Bamberg University (BERG Working Paper Series on Government and Growth), January 2009.

NIG-Levy process in asset price modeling: case of Estonian companies

Dean Teneng[1]

Abstract. As an asset is traded at fair value, its varying price trace an interesting trajectory reflecting in a general way the asset's value and underlying economic activities. These trajectory exhibit jumps, clustering and a host of other properties not usually captured by Gaussian based models. Levy processes offer the possibility of distinguishing jumps, diffusion, drift and the laxity to answer questions on frequency, continuity, etc. An important feature of normal inverse Gaussian-Levy (NIG-Levy) model is its path richness: it can model so many small jumps in a way that eliminates the need for a Gaussian component; hence, limitations arising from Gaussian based models are almost eliminated. Secondly, the characteristics listed above are reflected in the Levy triplet and are easily introduced in the modeling picture through estimated Levy parameters. Thirdly, knowledge of NIG-Levy parameters enables us to use NIG-Levy models as underlying asset price models for pricing financial derivatives.

We use the R open software to calculate Levy parameters for 12 Estonian companies and choose good NIG-Levy asset price models by the method proposed by Käärik and Umbleja (2011). We observe that not all financial data of Estonian companies trading on the Tallinn Stock Exchange between 01 Jan 2008 – 01 Jan 2012 can be effectively modeled by NIG-Levy process, despite having Levy parameters. Those positively modeled are recommended as underlying assets for pricing financial derivatives.

Keywords: NIG, goodness of fits test, fitting price process.

JEL Classification: C68, C29
AMS Classification: 60G51, 65C20

1 Introduction

As an asset is traded at fair value, its varying price trace an interesting trajectory reflecting in a general way the asset's value and underlying economic activities [4]. These trajectory exhibit jumps, clustering and a host of other properties not usually captured by Gaussian based models. Levy processes offer the possibility of distinguishing jumps diffusion, drift and the laxity to answer questions on frequency, continuity, etc. An important feature of normal inverse Gaussian-Levy (NIG-Levy) model is its path richness: it can model so many small jumps in a way that eliminates the need for a Gaussian component; hence, limitations arising from Gaussian based models are almost eliminated [2, 6, 7].

Brownian motion or Gaussian based models are known to have serious limitations like light tails, inability to effectively capture jumps, model stochastic volatility, clustering, and a host of others extensively discussed in [2, 3]. Particularly important is their inability to make a clear distinction between rare large jumps and small but frequent jumps; one of the motivations calling out for the use of more efficient functions to model asset prices. One may even guess that in a small economy like Estonia, asset prices may not very much reflect global trends like market data collected from big economies with uncountable number of factors affecting asset prices. In this light, there is a possibility to think local elections may be reflected more in the prices of assets than Iran's current turbulent relation with the west affecting the prices of oil on the world market. As this event based approach is the case, we need more responsive functions to reflect such trends than ordinary Gaussian based models.

We calculate NIG-Levy process parameters by maximum likelihood method and apply Käärik and Umbleja [5] proposed strategy for selecting good models. We observe that the asset price trajectories of two companies (Baltika and Ekpress Grupp) trading on the Tallinn stock exchange between 01 Jan 2008 – 01 Jan 2012 can be modeled by NIG-Levy process. Thus their stochastic properties are effectively described by NIG-Levy model and these price models can be used by investors in constructing optimal portfolios.

The following section answers the question why we need normal inverse Gaussian Levy process for modelling asset prices. It briefly touches on general characteristics of Levy processes before expanding with the NIG-

[1] University of Tartu/Institute of Mathematical Statistics, J.Liivi 2-518, 50409 Tartu, Estonia, teneng@ut.ee

Levy process. Analysis of data and selection of model is the subject of section three. Some useful graphs are appended in the appendix.

2 Why normal inverse Gaussian Levy process for modeling asset prices?

The normal inverse Gaussain Levy process is a member of the general class of Levy processes. By definition, a stochastic process X_t is a Levy process if has stationary and independent increments, and has stochastically independent paths. Mathematically speaking, a Levy process is a continuous time stochastic process $X = \{X_t : t \geq 0\}$ defined on the probability space (Ω, F, P) with the following basic properties:

1) $P(X_0 = 0) = 1$ i.e the process starts at zero;
2) $\forall\, s, t \geq 0, X_{s+t} - X_t$ is distributed as X_s i.e. stationary increments;
3) $\forall\, s, t \geq 0, X_{s+t} - X_t$ is independent of X_u, $s \leq t \leq u$ i.e. independent increments;
4) $t \to X_t$ is a.s. right continuous with left limits.

The above properties clearly describe closing price properties, and their study in the general framework of Levy processes reveals useful properties. Levy processes present us with models that eliminate most of the weaknesses of Gaussian models. The make clear distinctions between large and small jumps and are able to capture small jumps in a way that we do not need a Brownian motion component. Further, they present possibility of answering questions related to frequency, size, continuity as well as distinguishing between drift, Brownian motion and Poisson based components making each of these only a limiting case [6,7]. Associated with Levy processes are infinitely divisible probability distributions, with the possibility of skewed shapes and slow decaying tails that perfectly fit log-return data [1, 2, 6]. Some of these properties can be seen by applying the famous Levy-Ito decomposition theorem and arriving at the Levy-Khinchine representation of a general Levy process as can be seen clearly in [1, 6] and for the NIG-Levy case presented in equation (3) below.

NIG-Levy process definition and basic characteristics

NIG-Levy process with parameters $\alpha, \beta, \mu, \delta$ and condition $-\alpha < \beta < \alpha$ denoted $NIG(\alpha, \beta, \delta, \mu)$ can be defined as follows:

Consider a bivariate Brownian motion (u_t, v_t) starting at point $(u, 0)$ and having constant drift vector (β, γ) with $\gamma > 0$ and let z denote the time at which v_t hits the line $v = \delta > 0$ for the first time (u_t, v_t are assumed independent). Then letting $\alpha = \sqrt{\beta^2 + \gamma^2}$, the law of u_z is $NIG(\alpha, \beta, \delta, \mu)$ [1]. This distribution has probability density of the form

$$f_{NIG}(x : \alpha, \beta, \delta, \mu) = \frac{\alpha\delta}{\pi} e^{\left(\delta(\alpha^2-\beta^2)^{1/2} - \beta(x-\mu)\right)} \frac{K_1\left(\alpha(\delta^2 + (x-\mu)^2)^{1/2}\right)}{(\delta^2 + (x-\mu)^2)^{1/2}} \tag{1}$$

where $K_1(x) = \frac{x}{4}\int_0^\infty e^{\left(t + \frac{y^2}{4t}\right)} t^{-2}\, dt$ is a modified Bessel function of the third kind. $\delta > 0$ is scale, $\beta \geq 0$ is symmetry, μ is location and $\alpha > 0$ is tail heaviness. It has a moment generating function of the form

$$M(x : \alpha, \beta, \delta, \mu) = exp\left[\delta\left\{\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + x)^2}\right\} + x\mu\right], \tag{2}$$

mean = $\mu + \frac{\delta\beta}{\sqrt{\alpha^2-\beta^2}}$, variance = $\alpha^2\delta(\alpha^2 - \beta^2)^{-3/2}$, skewness = $3\beta\alpha^{-1}\delta^{-1}(\alpha^2 - \beta^2)^{-1/4}$ and kurtosis = $3\left(1 + \frac{\alpha^2 + 4\beta^2}{\delta\alpha^2\sqrt{\alpha^2-\beta^2}}\right)$ [6]. A useful property is the fact that if $x_1, x_2, x_3, \ldots, x_m$ are independent normal inverse Gaussian random numbers with common parameters α, β but having individual location and scale parameters u_i and $d_i (i = 1, \ldots, m)$, then $x_+ = x_1 + \cdots + x_m$ is again distributed according to a normal inverse Gaussian law, with parameters $(\alpha, \beta, \delta_+, u_+)$ [1]. Thus closing prices can be drawn from such a distribution in two ways. First, each closing price can be drawn from a different NIG-Levy distribution with individual location and scale parameters but all the corresponding NIG-Levy distributions have common α, β parameters. Second, all the closing prices can be drawn from the same NIG-Levy distribution with unique $(\alpha, \beta, \delta, \mu)$ parameters. We employ the second method in this work.

The characteristic function of $NIG(\alpha, \beta, \delta, \mu)$ can be represented as $E[e^{iuX_t}] = e^{t\psi(u)}$ giving us the Levy-Khinchine representation as

$$\psi(u) = \int_{|x|\geq 1} (1 - e^{iux}) f(x : \alpha, \beta, \delta)\, dx + \int_{|x|<1} (1 - e^{iux} - iux) f(x : \alpha, \beta, \delta)\, dx + iu\rho \tag{3}$$

where $f(x: \alpha, \beta, \delta) = \frac{\delta\alpha}{\pi|\mathrm{x}|}\exp(\beta x)K_1(\alpha|\mathrm{x}|)$ and $\rho = \frac{2\delta\alpha}{\pi|\mathrm{x}|}\int_0^1 \sinh(\beta x)\, K_1(\alpha x)$ and $\mu = 0$ here. The first term of $\psi(u)$ (equation 3) capture big jumps while the second term captures small jumps [1].

Considering that closing prices of assets can be drawn from a Levy process with defined characteristics, the above mentioned stylized facts not generally captured by Brownian models are incorporated in Levy models through the characteristic triplet. This in itself means that Levy models are more effective than pure Gaussian models. Secondly, determination of Levy parameters means we can represent asset price process with Levy model. It must be injected here that this says nothing about the effectiveness of the model, only that it is better than pure Gaussian models. Thus, we still need goodness of fits test and other strategies to select best model.

3 Selecting best fit asset price model

We used the method proposed by Käärik and Umbleja [5] for selecting best asset price models. Below, the method is outlined.

Käärik and Umbleja proposed method for selecting best models

1. choose a suitable class of distributions (using general or prior information about the specific data) ;
2. estimate the parameters (by finding maximum likelihoods);
3. estimate goodness of fit;
   a) visual estimation,
   b) classical goodness-of-fit tests (Kolmogorov-Smirnov, chi-squared with equiprobable classes),
   c) probability or quantile-quantile plots.

This method was proposed as a general method in the fitting of insurance claim data to distributions. The distributions they used had tails, with some having really heavy-tails and even belonging to the class of subexponential distributions. But the proposed method is general. It turns out that NIG-Levy process has semi heavy tails, thus a good candidate for using the model. The subject of proving whether NIG-Levy process is subexponential or not is still being investigated, even though there is considerable literature showing that certain aspects of the NIG-Levy process can be restricted to fulfill subexponentiality. As well, analysis of data for different companies revealed the presence of tails. These and other observable properties are discussed below.

Company	Alpha (α)	Beta (β)	Delta (δ)	Mu (μ)
Arco Vara	468.90	468.86	0.03	0.02
Baltika	7.06	6.62	0.22	0.52
Ekpress Grupp	2.68	2.15	0.49	0.85
Harju Elekter	3.20	-2.07	0.72	2.95
Tallinna Kaubamaja	145.95	-145.06	0.55	9.75
Tallinna Vesi	107.80	106.86	0.85	2.92
Trigon	2953.24	2952	0.01	0.21
Nordecon	4.78	4.39	0.38	0.86
Viisnurk	208.64	-203.95	0.32	2.70
Olympic Entertainment Grupp	1107.89	1106.56	0.07	-0.03
Silvano Fashion Grupp	35.31	-29.589	8.11	14.43
Tallink	1588.86	356.46	49.54	-10.82

Table 1 Estimated NIG-Levy parameters for different companies trading on Tallinn stock exchange between 01/01/2008 and 01/01/2012

Analysis and selecting best models

To implement step one; we analyzed the kurtoses and skews of different companies. The skews of four companies (Harju Elekter, Tallinna Kaubamaja, Trigon, and Viisnurk) were negative indicating left tails while the remaining were positive indicating right tails. Thus data told us to use distribution that can capture tails. This is very good as the NIG-Levy process can capture tails. Theoretically, this feature comes to light when asymptotic relations for Bessel functions are applied. Next, kurtoses were looked at. None of the companies had a zero kurtosis thus data was non-normal. Of the positive kurtoses, Baltika and Ekpress Grupp had highest peaks. Hence, data suggested using distribution that has peaks and can capture tails. Peaks are essentially captured by NIG-Levy process as these represent jumps of the process. Both large and small jumps are captured.

Parameters for NIG-Levy process were estimated using maximum likelihoods as suggested and results for different companies are displayed on Table 1.

Company	$\chi 2$ statistic	$\chi 2$ p-value	K-S D-value	K-S p-value	Skew	Kurtosis
Arco Vara	2251.60	P < 0.00001	0.23	p < 0.00001	0.38	-1.53
Baltika	1771.12	P < 0.00001	0.06	p = 0.05723	1.67	2.33
Ekpress Grupp	1194.24	P < 0.00001	0.07	P = 0.01198	1.70	2.53
Harju Elekter	1345.87	P < 0.00001	0.09	p = 0.00027	-0.82	-0.05
Tallinna Kaubamaja	919.71	P < 0.00001	0.94	P < 0.00001	-1.18	-1.21
Tallinna Vesi	1225.08	P < 0.00001	1	P < 0.00001	1.11	1.05
Trigon	927.70	P < 0.00001	0.62	P < 0.00001	-0.03	-1.48
Nordecon	2287.88	P < 0.00001	0.89	P < 0.00001	0.004	-0.92
Viisnurk	741.17	P < 0.00001	0.92	P < 0.00001	-0.52	-0.72
Olympic Entertainment Grupp	1661.61	P < 0.00001	1	P < 0.00001	0.42	-0.68
Silvano Fashion Grupp	1430.68	P < 0.00001	0.68	P < 0.00001	1.58	1.63
Tallink	937.89	P < 0.00001	0.86	P < 0.00001	1.37	0.77

Table 2 Chi-square, Kolmogorov –Smirnov test values and sample skew and kurtosis for different companies trading on Tallinn stock exchange between 01/01/2008 and 01/01/2012

We observed that chi-square test rejected the null hypothesis (theoretical distribution suits described data) for all the companies while Kolmogorov-Smirnov (KS) test had positive feedback for three companies (Baltika, Ekpress Grupp, and Harju Elekter). Visual estimation of the fitted density and log density plots (see appendix for plots) led to selecting four companies (Nordecon, Trigon, Baltika and Ekpress Grupp) as best candidates. Further visual estimation by studying Q-Q plots eliminated Trigon and Nordecon. We were thus left with Baltika and Ekpress Grupp as the only suitable candidates for NIG-Levy asset price model.

4 Conclusion

We have reviewed the essential properties of Levy processes that make them more suitable for asset price modelling than general Gaussian processes and applied the NIG-Levy process to modeling asset prices for assets traded on the Tallinn stock exchange between 01 January 2008 and 01 January 2012. Selecting best NIG-Levy asset price model using the strategy proposed by Käärik and Umbleja [2011], we concluded that prices of Baltika and Ekpress Grupp can be modeled with NIG-Levy process; thus can be used as underlying assets in pricing financial deriveatives.

Acknowledgements

Research was supported by Estonian Science Foundation Grant No.8802 and Estonian Doctoral School in Mathematics and Statistics.

References

[1] Barndorff-Nielsen, O.E.: Processes of the Normal Inverse Gaussian type. *Finance and Stochastics* **2** (1998), 41-68.

[2] Carr, P., and Wu, L.: Time-changed Levy processes and option pricing. *Journal of Financial Economics* **71** (2004) 113-141.

[3] Cont, R.: Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* **Volume 1** (2001) 223-236.

[4] Gentle, E.J., and Härdle, W.K.: Modelling Asset Prices. *SFB 649 Discussion Paper 2010 – 31*, accessed 03/03/2010 at http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2010-031.pdf

[5] Käärik, M., and Umbleja, M.: On claim size fitting and rough estimation of risk premiums based on Estonian traffic insurance example. *International Journal of Mathematical Models and Methods in Applied Sciences* **Issue 1 Volume 5** (2011) 17-24.

[6] Schoutens, W.: *Levy processes in Finance*. John Wiley & Sons Inc., New York, 2003.

[7] Teneng, D.: Path properties of Levy processes. In: *Proceedings of the First International Scientific Conference of Students and Young Scientists '11* (Lytvynenko, I. O., and Terletskyi, D. O., eds.). Theoretical and Applied Aspects of Cybernetics, Kyiv, 2011, 214-218.

Appendix

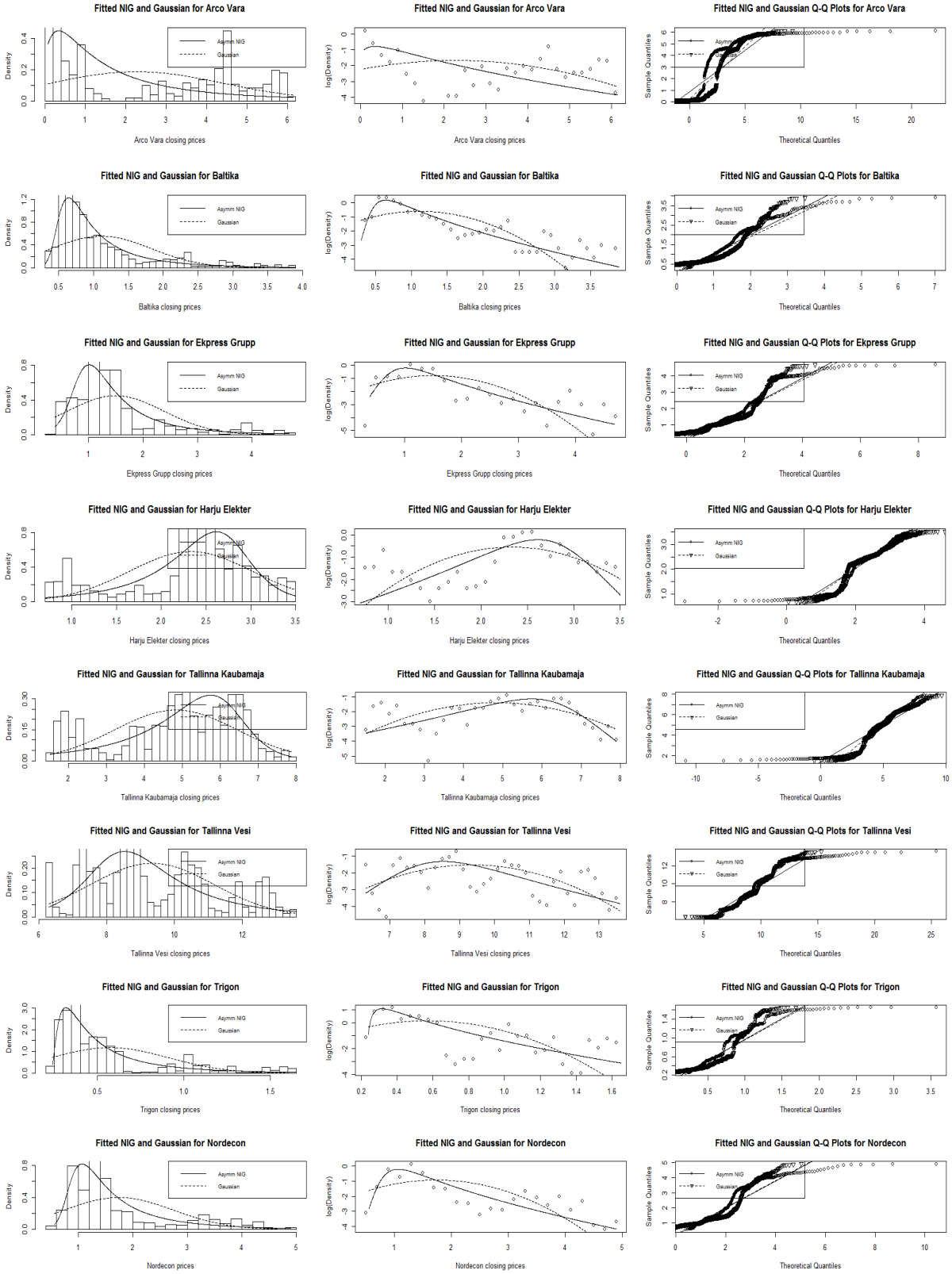


Figure 1 Plots of fitted NIG-Levy and Gaussian densities, log densities and Q-Q plots to the closing prices of 8 companies trading on the Tallinn stock exchange between 01/01/2008 and 01/01/2012

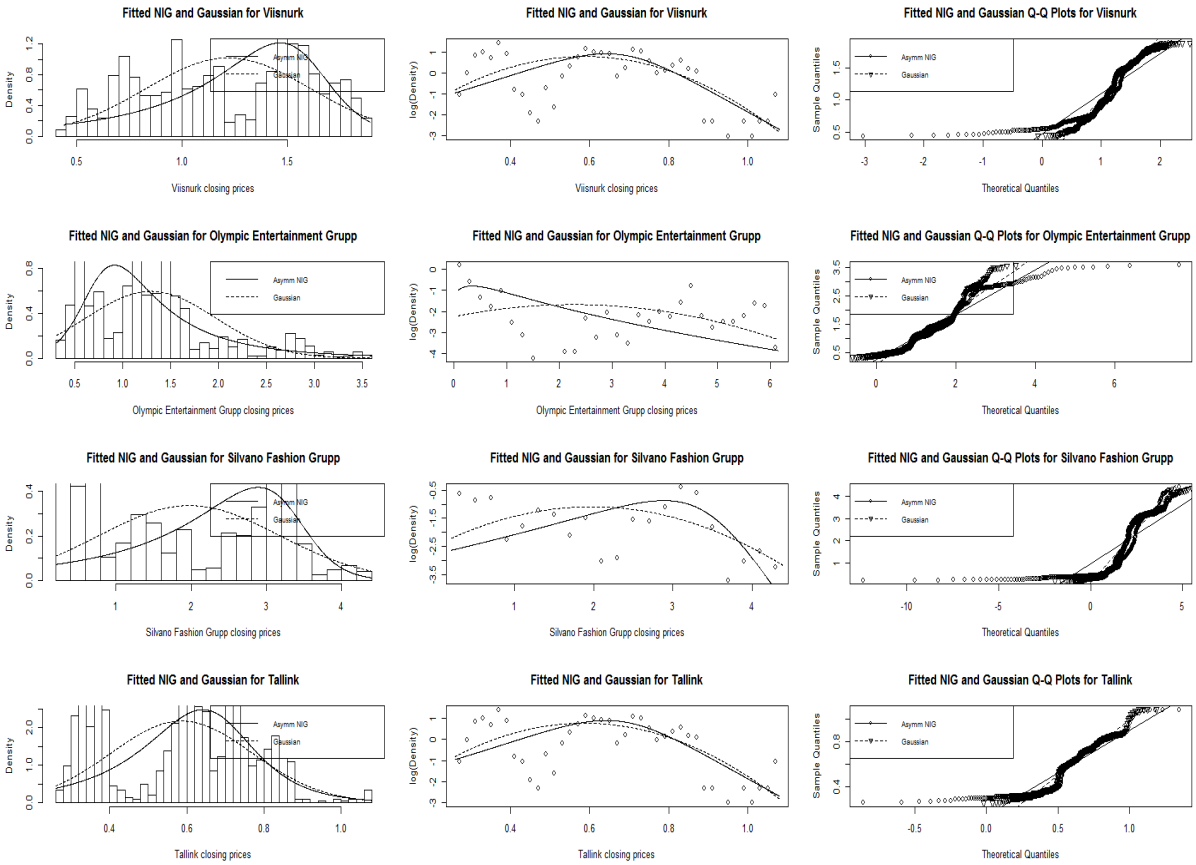


Figure 2 Plots of fitted NIG-Levy and Gaussian densities, Log densities and Q-Q plots to the closing prices of 4 companies trading on the Tallinn stock exchange between 01/01/2008 and 01/01/2012

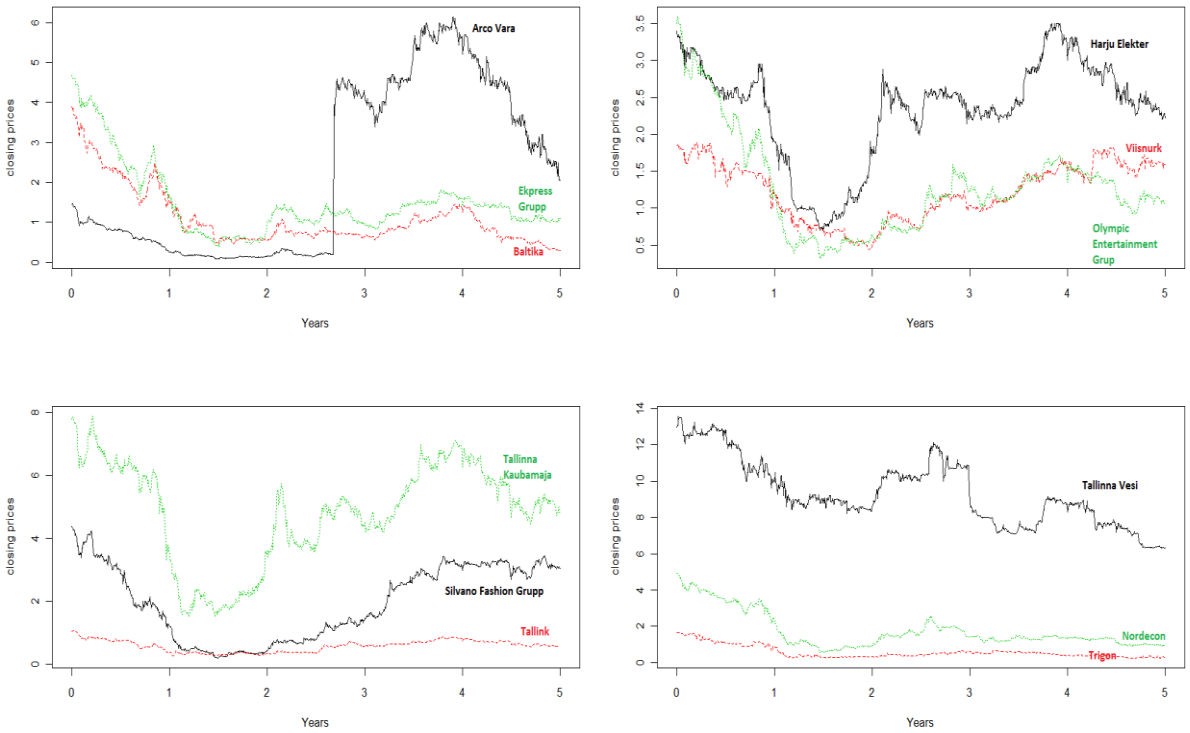


Figure 3 Plots of closing prices (thousands of euros) of 12 companies trading on the Tallinn stock exchange between 01/01/2008 and 01/01/2012

Some findings about risk estimation and backtesting at the world FX rate market

Tomáš Tichý[1]

Abstract. Financial markets are very sensitive to all kinds of risk. Immediately after any unexpected announcement the volatility of price returns is suddenly increased and market prices can potentially fall down. However, the announcement can influence prices of only some assets, while prices of others may remain relatively stable. It follows that a different risk type indicates a need for distinct methods of risk modelling, measuring and managing. In this paper we continue in our previous research and try to identify if there is any similarity in risk estimation model performance across particular world FX rate markets and provide the most important findings about the (dis)similarities with special attention to (former) transitional economies of Europe and Asia-Pacific regions. In particular, we apply VG and NIG models of marginal distribution for VaR calculation and backtesting and use JPY and USD currencies as examples of global FX rates with potentially low tights to the regional evolution. We show some interesting results about the impact of specific information arrival.

Keywords: Backtesting, FX rate market, information arrival, Lévy models.

JEL classification: G17
AMS classification: 91B30

1 Introduction

Financial markets are very sensitive to all kinds of risk. Immediately after any unexpected announcement the volatility of market returns is suddenly increased and market prices can potentially fall down. However, the announcement can influence prices of only some assets, while prices of others remain stable. The reason is that various assets are sensitive to distinct kinds of risk in a different way. It also follows that a different risk type indicates a need for distinct methods of risk modelling, measuring and managing.

During last years, we could observe many innovations at the market. Such events concerned all market segments, including production, services as well as financial services. These events sometimes accompanied by political changes and crises results into regular arrival of information that can make a change in the market equilibrium. It is natural that in order to model the behavior of market prices one needs very progressive tools. The contemporary research therefore concerns on advance models, possibly with jumps, that can model such features (see eg. [5] for comprehensive review of suitable models and references therein for more detailed information). Another line of research is focused on testing the model ability to estimate the risk measures in a sound way (see eg. [1] and references therein).

We already had the opportunity to study several advanced Lévy models based on the subordinator as a proxy to market activity and information arrival. For example, in [10] an important contribution of VG/NIG model with Student copula for the soundness of risk estimation was provided and subsequently. Next, in [6, 7] some further results on the model estimation via combined time span were provided. By contrast, [3] analyzed the impact of the security type (FX rates, single stocks) on the model performance, while in [4] the FX rates were analyzed in more details with special attention paid to former transitional markets of Central Europe.

Similarly to [4], the aim of this paper is to identify – on the basis of a risk model performance, including backtesting procedure – if there is any similarity among selected currencies or, to be more exact, their exchange rates with respect to Euro, and especially whether integration of new economies of Asia-Pacific

[1]Technical University Ostrava, Department of Finance, Sokolská 33, 701 21 Ostrava, Czech Republic, tomas.tichy@vsb.cz.

region as well as Europe to the global market implies some similarities in risk estimation failures or if they still behave differently.

We proceed as follows. In the following section we briefly review the problem of market risk modeling with special focus at modeling via subordinated Lévy models and risk model validation (in line with [6]. Next, in Section 3 we describe the data used in this study and finally, in Section 4, particular models are applied in order to examine the (dis)similarities of given FX rates.

2 Risk modelling tools

Assuming a random variable X following a Gaussian distribution,VaR over a time length Δt with a significance α (i.e. with confidence $1-\alpha$) can be obtained as follows:

$$VaR_X(\Delta t, \alpha) = -F_X^{-1}(\alpha) = -\mu_X(\Delta t) - \sigma_X(\Delta t) F_{\mathcal{N}}^{-1}(\alpha). \tag{1}$$

Here, $-F_X^{-1}(\alpha)$ denotes the inverse function to the distribution functionof random variable X for α (similarly, $F_{\mathcal{N}}$ is used for standard normal (Gaussian) distribution). It follows, that:

$$Pr\left(X < -VaR_X(\Delta t, \alpha)\right) = \alpha. \tag{2}$$

However, it is a rare case that random variable, eg. a return of financial asset, follows Gaussian distribution. Usually, we have to select a distribution with some additional parameters, so that we can control even higher moments of the distribution. In that case, it can be inevitable to run a Monte Carlo simulation procedure to obtain VaR as an estimate to $-F_X^{-1}(\alpha)$. Several useful models belongs to the family of Lévy processes or subordinated Lévy processes in particular (see eg. [5]).

Let us define a stochastic process $\mathcal{Z}(t; \mu, \sigma)$, which is a Wiener process. As long as $\mu = 0$ and $\sigma = 1$ its increment within infinitesimal time length dt can be expressed as:

$$dZ = \varepsilon\sqrt{dt}, \quad \varepsilon \in \mathcal{N}[0,1], \tag{3}$$

where $\mathcal{N}[0,1]$ denotes Gaussian distribution with zero mean and unit variance. Then, a subordinated Lévy model can be defined as a Brownian motiondriven by another Lévy process $\ell(t)$ with unit mean and positive variance κ. The only restriction for such a driving process is that it is non-decreasing on a given interval and has bounded variation. Hence, we replace standard time t in

$$X(t; \mu, \sigma) = \mu dt + \sigma\mathcal{Z}(t) \tag{4}$$

by its function $\ell(t)$:

$$X(\ell(t); \theta, \vartheta) = \theta\ell(t) + \vartheta Z(\ell(t)) = \theta\ell(t) + \vartheta\varepsilon\sqrt{\ell(t)}. \tag{5}$$

Due to its simplicity (tempered stable subordinators with known density function in the closed form), the most suitable models seem to be either the variance gamma model – the overall process is driven by gamma process from gamma distribution with shape a and scale b depending solely on variance κ, $G[a,b]$, or normal inverse Gaussian model – the subordinator is defined by inverse Gaussian model based on inverse Gaussian distribution, $IG[a,b]$. In this paper, we will apply the latter. Therefore, the risk of a single position can be estimated by evaluating the log-return model:

$$X(I(t;\kappa); \theta, \vartheta) = \mu t + \theta(I(t) - t) + \vartheta Z(I(t)) = \mu t + \theta(I(t) - t) + \vartheta\varepsilon\sqrt{I(t)}, \tag{6}$$

where μ is average return (long-term drift), which is fitted by deducing θt from the original model, and I is an inverse Gaussian process independent of ε.

Since VaR is often calculated for portfolios, the models for marginal distribution of asset returns as the one in (6) must be joined. A very useful way is to utilize copula functions. Assuming for simplicity two potentially dependent random variables with marginal distribution functions F_X, F_Y and joint distribution function $F_{X,Y}$, we get, according to Sklar's theorem (see [9] for more details):

$$F_{X,Y}(x,y) = \mathcal{C}(F_X(x), F_Y(y)). \tag{7}$$

Thus, being equipped by copula function and marginal distribution function, the joined evolution can be modeled easily.

The ability of (market) risk models to estimate the risk exposure soundly is commonly assessed by the so called backtesting procedure. Within the backtesting procedure on a given time series $\{1, 2, \ldots, T\}$, two situations can arise – the loss L is higher than its estimation or lower (from the stochastic point of view, the equality shouldn't arise). While the former case is denoted by 1 as an exception, the latter one is denoted by zero:

$$I_X(t+1,\alpha) = \begin{cases} 1 & \text{if} \quad L_X(t,t+1) > VaR_X(t,t+1;\alpha) \\ 0 & \text{if} \quad L_X(t,t+1) \leq VaR_X(t,t+1;\alpha). \end{cases} \tag{8}$$

On the sequence $\{I_X(t+1,\alpha)\}_{t=1+m}^{T-1}$, where m is a number of data (days) needed for the initial estimation, it can be tested whether the number of *ones* (exceptions) corresponds with the assumption, ie. $\alpha \times n$ (where $n = T - 1 - m$), whether the estimation is valid either unconditionally or conditionally, whether bunching is present, etc. Generally the most simple way is to compare the true number of exceptions to the assumption about them [8]. A one step further is to evaluate the one-step dependency of exceptions in line with [2]. The review of some further techniques can be found eg. in [1].

3 Data

The data set we consider in this study comprises of daily prices of several currencies from Europe and Asia-Pacific area in terms of Euro starting in January 2, 2001 to December 31, 2010. Particular currencies are evident from the first column of Table 1 – we consider currencies of (i) Indonesia, South Korea, Malaysia, Philippines, Thailand, (ii) Japan, (iii) the Czech Republic, Hungary, Poland and USA. Over given period we could collect about 2,520 observations of log-returns of each FX rate. It allows us to leave about 2 years for initial parameter estimation of particular models and then realize a backtesting procedure for 2,000 subsequent days, ie. observed number of exceptions can be easily transformed into probability. In the same table we provide basic descriptive statistics of particular FX rates: mean, standard deviation (both per annum), skewness and kurtosis.

It is evident that only three FX rates were appreciating with respect to EUR – and only one of them was appreciating strongly. By contrast, USD was the only FX rate strongly depreciating. Concerning the standard deviations, values close to 10% p.a. were mostly recorded. There are just two exceptions, one in each direction – again CZK (only 6.68%), and IDR (16.51%). According to observed values of skewness and kurtosis, it seems that there are two or three FX rates which are relatively close to Gaussian distribution – USD, MYR and THB. For the others, the kurtosis is slightly above 10 and in one case even above 40 (PHP) – for the latter we could observe also quite high skewness, which is surprisingly positive.

FX rate	Mean	St.dev.	Skewness	Kurtosis
IDR	−1.78%	16.51%	0.181	13.7446
KRW	−0.49%	13.14%	−0.1524	12.2002
MYR	0.77%	9.89%	−0.0992	5.3699
PHP	−1.55%	11.92%	2.059	43.4126
THB	0.37%	11.02%	−0.1857	5.8233
JPY	−0.16%	12.41%	0.1799	8.9864
CZK	3.32%	6.68%	0.0314	10.2736
HUF	−0.47%	10.04%	−0.6868	14.0017
PLN	−0.31%	11.08%	−0.3377	14.5372
USD	−3.59%	10.44%	0.0331	5.8854

Table 1 Basic descriptive statistics of FX rates (daily log-returns)

4 Results

In this section, we first estimate the probability distribution and VaR for particular FX rates ex-post and then we realize backtesting procedure to analyze potential similarities in model failures. We will consider VG and NIG model which will be compared to assumption of the standard market model (Brownian motion, BM).

Ex-post modeling

First, the VG and NIG model parameters will be obtained by the method of moments using all available data (log-returns). Next, we will use these parameters to estimate basic descriptive statistics for both models (VG, NIG) via a Monte Carlo simulation with 1,000,000 independent scenarios and compare them to the real observations depicted in Table 1.

The parameter to fit the volatility (ϑ) is the same for both models. Concerning the other parameters, we can see that in comparison to the NIG model, the VG model can lead to slightly lower values of θ (in absolute terms), but higher values of κ. It is also confirmed that negative skewness implies a negative value of θ, although for example for PHP we can observe that the skewness parameter is insignificant in relation to its huge kurtosis.

Since both models are defined on the basis of a subordinator following a different distribution (either gamma or inverse Gaussian), it is natural that the parameters will differ. Nevertheless, since the fitting was based on the method of moments, estimated moments for both models should be (approximately) the same and should approach the empirical levels. The results for the VG and NIG models are therefor much better than those, we might obtain via Gaussian distribution of BM. Still, however, we can observe some error in estimated kurtosis (about 5%). It is probably given by the fact, that the rare events are difficult to much – unexpected jumps in log-returns that are modeled according to a given specification of parameters for a given (and finite) number of scenarios in fact results into non-smooth approximation of the log-distribution function in the tails. On the other hand, it is difficult to guess weather the realization is far from reality or not, since also the number of market data is finite.

From the point of view of risk management, modeling of quantiles (VaR) is even more important than matching the moments of the distribution. It is also obvious that observed deficiency in kurtosis estimation will result into differences in VaR estimation. Moreover, we should note that with a given number of data, VaR at 0.1% probability level will be obtained as the third worst result, while within the simulation it is 1,000th worst result.

It is therefore much more important to check if there are some differences between both models and among particular FX rates – generally, we cannot identify any significant differences between the VG and NIG models, as well as among particular FX rates in both regions. Of course, sometimes the results for a given model/FX rate differ more than for another combination, but it cannot lead to any strong conclusion about the model/FX rate dissimilarity. We therefore proceed to backtesting procedure in the next section.

4.1 Backtesting

In the following text we will concentrate on the backtesting results of single positions in particular FX rates assuming three different significance levels (0.1%, 1% and 5%) for VaR estimation from the point of view of either the long position or the short position, which will be accompanied by the median. We will consider a standard market model (Brownian motion, BM) and the VG and NIG models. In order to estimate the parameters of these models we will use returns over either 250, 500 or 1,000 preceding days, that is about 1, 2 and 4 years, respectively.

We start on the same day for both cases so that the total number of loss observations is always identical. In particular, we start on day t and use either 250 or 500 or 1,000 preceding returns to estimate the next day's VaR ($t + 1$) for particular significance levels – it also indicated that the initial estimate is done on day $t = 501$. Next it is compared to the actual observed loss (return with minus sign) and either one or zero is recorded. This procedure is repeated on a moving window basis until we reach the last day. (The total number of exceptions and the values of the test statistics for all combinations can be made available upon request.)

Starting with the standard market model (BM) and 250-day window, we can clearly see that the model does not allow us to get generally acceptable results for the most frequent significances of 0.1% and 1% (for both positions, short and long, and all FX rates). It is interesting to see that while for JPY we get acceptable results only assuming long position, with another FX rates (eg. THB, MYR), we can get quite good results when short position is assumed. One might suggest that the reason is the modest value of positive skewness – but we could observe the same value even for IDR. In this case, however, there is quite high kurtosis. Thus, it is evident that BM model works well if low kurtosis is combined with

positive skewness (assuming long position) or negative skewness (assuming short position). By contrast, if the kurtosis is about 8 or higher, the BM model is suitable neither for long nor short position risk modeling. It is very important to keep these observations in mind when making decision about model suitability just on the basis of one-tail estimates.

Moreover, we also observe that the results for VaR5% and the median are always satisfactory. It is obvious, since the 'common' values should not be influenced by higher moments of the distribution (skewness and kurtosis). It is also interesting to note that JPY, the only FX rate with positive skewness and modest kurtosis results into lower number of exceptions on the left than on the right even for VaR5%.

Concerning the results for extended time window used for estimation, since mean and standard deviation are rather of a short memory, the impact should be rather negative. We can check it in the tables – the number of exceptions increased for all cases, though there is no important impact from the statistical point of view. The cases, in which the model was accepted, result to acceptance again, and vice versa.

Now, we can proceed to the more advanced models, VG and NIG. Since both these alternatives should allow us to fit the observed skewness as well as kurtosis well, the results should be much better. Obviously, for some FX rates the results are better than for the others, but generally we can say that both models provide us acceptable results for all considered probabilities and FX rates.

Note however, that a return volatility is a rather short-memory type measure. Finally, we thus try to combine two different window lengths for the estimation – we provide these results in Table 4. While the first two moments are estimated on a 250-day basis, in order to get the skewness and kurtosis we will use multiyear window. Moreover, the underlying data will be normalized to get zero mean and unit variance over a one year window. This will allow us to eliminate potential error in the volatility estimation, but to get a clearer picture of the impact of rare events on exceptions at particular markets. Note also that the volatility of particular FX rates was significantly different (see Table 1) – thus the normalization will also help us to overcome this feature.

Let us return to the analysis of exceptions. Initially, the model performance was assessed according to the Kupiec test, which is based only on the relative number of exceptions. It is evident that the subordinated Lvy model (VG and NIG) works very well for most of the other FX rates, except some special cases, regardless of the group. However, a high quality model should moreover not lead to any bunching, i.e., clusters of exceptions during subsequent days. Unfortunately, such testing requires relatively long data series and a higher number of observed exceptions. It is therefore not very suitable for far tails analysis, i.e., if we observe only two or three exceptions, any conclusion about potential bunching is not reliable. In our case, we can therefore test the exceptions series on bunching only for a 5% VaR and sometimes also 1% VaR. For some currencies, clusters can be identified when the long position is considered, for others they are present only in the case of the short position.

It would also be interesting to check what happens, if there is an exception in the risk estimation on a given FX rate – i.e., can the exceptions be observed on the same day for all FX rates? The answer can be found in Tables 2 and 3, where the relative conditional exceptions of single position VaR5% according to the NIG model for the first/last half of the days are depicted. For example, if the risk model for JPY fails, then the probability that the model for THB will also fails is either 0.25 or 0.51 during the first/last half, respectively. Note also that the probability of the reverse (the JPY model fails conditionally on a SGD model failure) is totally different, which is obviously given by the different number of exceptions for both FX rates. However, if we consider VaR5%, we get almost identical results.

	First half					Second half				
FX rate	JPY	CZK	HUF	PLN	USD	JPY	CZK	HUF	PLN	USD
JPY	1	0.16	0.07	0.06	0.33	1	0.1	0.02	0.03	0.38
CZK	0.15	1	0.28	0.32	0.05	0.08	1	0.29	0.33	0.1
HUF	0.11	0.44	1	0.36	0.07	0.01	0.29	1	0.43	0.06
PLN	0.06	0.35	0.25	1	0.1	0.03	0.41	0.53	1	0.05
USD	0.3	0.05	0.04	0.09	1	0.34	0.1	0.07	0.04	1

Table 2 Conditional exceptions of single position VaR5% according to NIG model for first/last 1,000 days (probability for European FX rates)

	First half						Second half					
FX rate	JPY	IDR	KRW	MYR	PHP	THB	JPY	IDR	KRW	MYR	PHP	THB
JPY	1	0.18	0.26	0.21	0.18	0.25	1	0.26	0.44	0.44	0.39	0.51
IDR	0.24	1	0.38	0.58	0.5	0.32	0.25	1	0.31	0.45	0.34	0.42
KRW	0.41	0.44	1	0.6	0.48	0.36	0.43	0.32	1	0.49	0.43	0.55
MYR	0.32	0.66	0.59	1	0.73	0.45	0.33	0.37	0.38	1	0.54	0.58
PHP	0.27	0.56	0.47	0.72	1	0.38	0.42	0.38	0.46	0.75	1	0.64
THB	0.35	0.34	0.33	0.42	0.36	1	0.47	0.41	0.52	0.71	0.57	1

Table 3 Conditional exceptions of single position VaR5% according to NIG model for first/last 1,000 days (probability for Asia-Pacific FX rates)

5 Conclusion

We have provided important results on market risk model performance in Asia-Pacific FX rate markets. The results may be interesting for financial policy evaluation and may also help financial institutions in the internal risk management process. An interesting extension of the analysis would be to model the FX rates as a portfolio and study the impact of particular FX rates.

Acknowledgements

This work was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070). The research of the second author was also supported by SGS project of VSB-TU Ostrava under No. SP2012/2.

References

[1] Berkowitz, J., Christoffersen, P.F., Pelletier, D.: Evaluating Value-at-Risk Models with Desk-Level Data, *Management Science* **57**: 2213–2227, 2011.

[2] Christoffersen, P.F.: Evaluating Interval Forecasts, *International Economic Rev.* **39**, 841–862, 1998.

[3] Cielepová, G., Tichý, T.: The implication of the security type for efficient risk measuring, *Actual Problems of Economics* **2** (12): 144–151, 2011.

[4] Cielepová, G., Tichý, T.: Interesting findings about risk estimation and backtesting at European FX rate market. In: R. Matoušek, D. Stavárek. *Financial Integration in the European Union*: 189–207. London: Routledge, 2012.

[5] Cont, R., Tankov, P.: *Financial Modelling with Jump Processes*, 2e. Chapman & Hall/CRC press, 2010.

[6] Kresta, A., Tichý, T.: International equity risk modeling by NIG model. In *Proceedings of Mathematical Methods in Economics*, Praha: Professional Publishing, 401–406, 2011.

[7] Kresta, A., Tichý, T.: International Equity Portfolio Risk Modeling: The case of NIG model and ordinary copula functions. *Finance a uver – Czech Journal of Economics and Finance* **61** (2): 141–151, 2012.

[8] Kupiec, P.: Techniques for verifying the accuracy of risk management models. *Journal of Derivatives* **3**, 73–84, 1995.

[9] Nelsen, R. B.: *An Introduction to Copulas*. 2nd ed. Springer, 2006.

[10] Tichý, T.: Posouzení odhadu měnového rizika portfolia pomocí Lévyho modelu. *Politická ekonomie* **58** (4): 504–521, 2010.

The use of the genetic algorithm for the upper bound calculation of the vehicle assignment problem

Ľubomír Toman[1]

Abstract. Evacuation is often used mean which serves to protect the population in case of emergency. In terms of operation research, the evacuation vehicle assignment problem (VAP) comprises the key part of the evacuation plan design problem. VAP represents nonlinear problem. This problem is solved by iterative method which is commonly used in fuzzy optimization to handle a nonlinearity model. With using this method in VAP, we fix the total evacuation time in order to obtain reduced vehicle assignment problem (RVAP) model. RVAP represents the hard combinatorial problem, which is computationally demanding. The branch and bound method is used for solving RVAP. The efficiency of this method is determined, among other things, by the method which serves to obtain the upper bound of the optimal solution. In this paper, we propose the use of the genetic algorithm to obtain the upper bound instead of the simply rounding heuristic. Numerical experiments were performed and the results draw the comparison between both of methods and illustrate the effectiveness/ineffectiveness of the proposed method which serves to obtain the upper bound of the optimal RVAP solution.

Keywords: genetic algorithm, branch and bound method, upper bound, nonlinear problem, evacuation.

JEL Classification: C61
AMS Classification: 90C27, 90C57

1 Introduction

Evacuation is a mean which serves for efficient protection of human lives and health. It is more common than many people realize. Hundreds of times each year, transportation and industrial accidents release harmful substances, forcing thousands of people to leave their homes [2]. The natural disasters are reasons which force people to be evacuated as well.

Evacuation belongs to the operational control of public service systems. We need to have an evacuation plan for efficient perform of evacuation. This plan allows us to evacuate in minimal time by vehicles people from endangered dwelling places to the safety places which let us named *refuges*. The target in evacuation plan design is to determine the route for each vehicle used in evacuation in order to every endangered inhabitant is evacuated to some refuge in minimal time. Total evacuation time is a time interval which starts when the vehicles depart from fleets and which stops when each inhabitant is evacuated into some refuge.

The evacuation plan design problem consists of following phases. In the first phase, it must be determined possible sets of refuges, fleets and endangered dwelling places for a particular emergency. Second phase divides endangered dwelling places into smaller part which have lesser number of inhabitants. The parts, which can be evacuated independently, are subsequently assigned to the refuges [5]. Let us name these parts as *municipalities*. Final phase of evacuation plan design problem consist in the route determination for each vehicle which is used for evacuation.

2 Vehicle assignment problem

After the first and second phase of evacuation plan design problem is done, we have input data for final phase and can formulate the *vehicle assignment problem* (VAP) [8].

There exists a set of municipalities J which are endangered by some threat. Each municipality $j \in J$ has b_j inhabitants who must be evacuated to the predetermined refuge. The municipality is pre-assigned to the refuge in advance. The set I is the set of homogenous fleets. Each fleet $i \in I$ has N_i vehicles with the same capacity. This capacity is given by a number of people who can be transported in this vehicle simultaneously. The target of the VAP is to assign appropriate number of vehicles from the fleets to the municipalities so that the evacuation can

[1] University of Žilina, Faculty of Management Science and Informatics, Department of Transportation Networks, Univerzitná 1, 01026 Žilina, Slovakia, lubomir.toman@fri.uniza.sk.

be performed in minimal time. Moreover, we assume that each vehicle can be assigned to one municipality at most. Considering this presumption, when the vehicle is assigned, the route of this vehicle is set as well.

2.1 Iterative method

The target of the VAP is to assign appropriate number of vehicles from the fleets to the municipalities so that the evacuation can be performed in *minimal* time. This problem is hard to solve. The results [3], [6] shows that better way to solve VAP is to use iterative method where in each iteration the reduced VAP (RVAP) is solved. The target of the RVAP is to assign appropriate number of vehicles from the fleets to the municipalities so that the evacuation can be performed in the *predetermined* time which is denoted with symbol T^{max}. RVAP is a decision problem. We are only interested in whether the feasible solution exists or not. If we are able to perform evacuation into the given time T^{max} so we decrease this time which represents the upper bound of the time of the optimal VAP solution and repeatedly solve the problem with the new value of T^{max}. If we are not able to perform evacuation in the given time then the time T^{max} +1 represents the lower bound of the time of the optimal VAP solution. The resultant solution is that feasible solution which was obtained for the lowest time T^{max}.

2.2 Reduced vehicle assignment problem

The reduced vehicle assignment problem is assignment problem where we try to assign appropriate number of vehicles from the fleets to the municipalities so that the evacuation can be performed in the *predetermined* time. Let the symbol q_{ij} denotes the number of vehicles from fleet i assigned for evacuation of the municipality j. Based on the predetermined time T^{max} and the travel times among fleets, municipalities and refuges we can calculate the values of coefficients a_{ij} for each couple i, j where $i \in I, j \in J$. Such coefficient a_{ij} represents the *evacuation capacity*, i.e., the number of people who can be evacuated from the municipality j by one vehicle from the fleet i into the time T^{max}. If a_{ij} has positive value then the vehicle from the fleet i can be used for evacuation of the municipality j, i.e. the municipality j is reachable from the fleet i in the time T^{max}. Let the symbol $J(i)$ denotes the set of municipalities $j \in J$ which are reachable from the fleet i and the symbol $I(j)$ denotes the set of fleets $i \in I$ which the municipality j is reachable from. The target of decision RVAP is to find a feasible solution which satisfies the constraints (1)-(3) or to prove that such solution does not exist.

$$\sum_{j \in J(i)} q_{ij} \leq N_i \quad for\ i \in I \tag{1}$$

$$\sum_{i \in I(j)} a_{ij} q_{ij} \geq b_j \quad for\ j \in J \tag{2}$$

$$q_{ij} \in Z_0^+ \quad for\ i \in I,\ j \in J(i) \tag{3}$$

The constraints (1) ensure that we use for evacuation only vehicles that the fleets contain. The constraints (2) ensure to every inhabitant from every municipality is evacuated. The **Figure 1** shows the graphic model of RVAP.

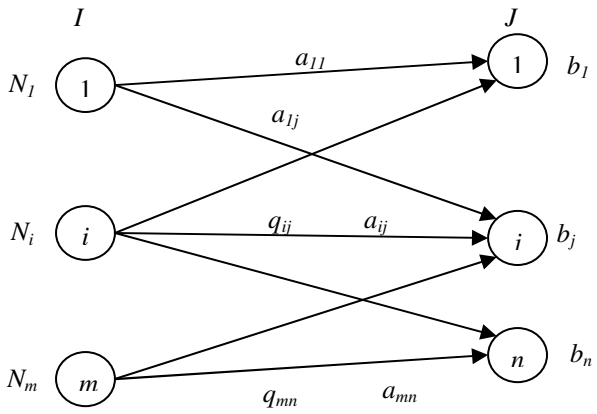


Figure 1 The graphic model of RVAP

3 Branch and bound method

RVAP represents hard combinatorial optimization problem. In this problem we are interested only if some solution exists which holds the constraints (1)-(3). This decision problem is transformed to the minimization one [10]. This minimization problem is solved by branch and bound (BB) method. In order to obtain the particular algorithm of the BB method the following parts must be set [4]:

1. searching tree scheme
2. way of branching
3. method for the lower bound calculation
4. method for the upper bound calculation

Efficiency of the BB method depends not only on the particular used parts but on the right combination of the used parts. In this paper we put emphasis on the methods for the upper bound calculation of RVAP. In the original approach the special rounding heuristic was used for this purpose. We propose to incorporate this heuristics into the genetic algorithm (GA) in order to obtain better results.

3.1 Rounding heuristics

In the original approach the special rounding heuristic is used for the upper bound calculation of RVAP. This heuristics uses the optimal solution of the LP relaxed minimization RVAP (or also the feasible solution of LP relaxed (1)-(3)) to obtain (whether feasible or infeasible) integer solution. In such LP relaxed solution every municipality $j \in J$ is satisfied. This heuristics works in two steps.

First step: For each pair i, j, where $a_{ij} > 0$ and the variable q_{ij} has noninteger value, round down this value if the municipality j remains satisfied. Otherwise, round up this value if the fleet i has enough vehicles. Otherwise, round down this value even if the municipality j becomes unsatisfied. Integer solution becomes temporarily infeasible. Note that if the value of variable q_{ij} is rounded down, we spare vehicles in the fleet i but decrease satisfaction of the municipality j and vice versa.

In the second step, we identify the unsatisfied municipalities and try to satisfy them with using spared vehicles from fleets. As first, we try to satisfy such municipality j where the value $I(j)$ is minimal. A feasible solution of (1)-(3) exists, if every municipality from the set J is satisfied after this procedure.

3.2 Genetic algorithm

The special rounding heuristics, which was initiated in 3.1, works quickly and therefore it is available to use this heuristics as a part of the BB method, where big number of the searching tree nodes must be processed. Efficiency of this heuristics depends on the order of processing the pairs i and j (i.e., on the order which the variables q_{ij} are rounded in). Unfortunately, we are not able to predetermine which order will lead to the best results. Therefore, it may happen that the potential of the heuristics won't be used.

We try to eliminate this disadvantage by the use of the genetic algorithm [12]. We use one of the features of the GA which allows distinguishing the genotype and the phenotype [7]. In our case, the genotype (or chromosome) includes the order which the couples i, j are processed in (i.e., the order which the variables q_{ij} are rounded in). The gene is the subscript of the particular fleet or municipality. The genotype consists of a fleet part and municipality part as the **Figure 2** shows. The fleet part contains the order which the fleets $i \in I$ are processed in. Analogically, the municipality part contains the order which the municipalities $j \in J$ are processed in.

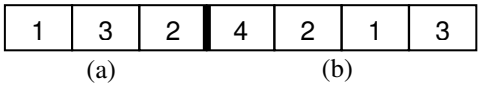


Figure 2 Genotype: (a) fleet part and (b) municipality part

The genotype on the **Figure 2** means that the rounding heuristics will round the values of the variables q_{ij} for these couples (i, j): (1, 4), (1, 2), (1, 1), (1, 3), (3, 4), (3, 2), … , (2, 1), and (2,3) respectively. Both of the genotype parts (the fleet and municipality part) have the same structure and they are processed in the same way. But each part is processed individually, e.g., to crossover genotype, we separately crossover parents fleet part and separately parents municipality part and consecutively join these offspring parts to create an offspring. In the following, we confine ourselves only to the fleet part of the genotype. Note that the same operations are applicable to the municipality part as well.

Let us assume the set I has m fleets. The fleet part contains permutation $\langle \pi_1, \ldots, \pi_m \rangle$ of fleet's subscripts 1, … , m. This permutation represents the order, which the fleets are processed in, in rounding process. We used

the partial mapped crossover (PMX) [7] scheme for crossover of parents. This crossover scheme holds a permutations of the subscripts 1, … , m in the created offsprings. The used operation of mutation changes several times positions of two randomly selected genes of the genotype part. The number of changes is a method parameter. In this approach the strategy of crossover-AND-mutation is used. The crossover of parents is carried out if the pseudorandomly generated number $r \in \langle 0; 1 \rangle$ is less then the crossover rate $\chi \in \langle 0; 1 \rangle$. The mutation of offsprings is carried out if the pseudorandomly generated number $r \in \langle 0; 1 \rangle$ is less then the mutation rate $\mu \in \langle 0; 1 \rangle$. The creation of the offspring genotype consists of the separate crossover and mutation of the fleet parent genotype and municipality parent genotype.

To use the selection it is need to evaluate the fitness of the individuals. When the rounding process of values q_{ij} is done, there are only two cases. First, the optimal solution of minimization RVAP was found. In the second case some of the municipalities are not fully satisfied, i.e., some portion of inhabitants was not evacuated from these municipalities. Number of such inhabitants makes the fitness of individuals. The lesser this number is, the better the fitness is. The rounding heuristics is the transformation which is used for genotype-phenotype mapping.

For the selection of individuals for parenthood we used roulette-wheel selection with ranking (RWS+R). One of the best advantages of the selection with ranking is a simple computation which can be done in $O(1)$ time [7]. We sort the individuals in ascending order of fitness. We have sorted sequence of individual $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ where n is the number of individuals. The probability of selecting the individual $\boldsymbol{x}_k$ is $p(k) = k/N$, where N is computed according to (4). When the pseudorandom number $r \in \langle 0; 1 \rangle$ is generated such kth individuals is selected for parenthood where (5) holds for the minimal k. Then the number k can be easily and quickly computed according to (6).

$$N = \sum_{i=1}^{n} i \tag{4}$$

$$\sum_{i=1}^{k} p(i) = \sum_{i=1}^{k} \frac{i}{N} \geq r \tag{5}$$

$$k = \left\lceil \frac{-1 + \sqrt{1 + 8Nr}}{2} \right\rceil \tag{6}$$

After the selection of individuals for parenthood, crossover, mutation and evaluating process is done, the better offspring is inserted into the set of offsprings. When the set of offsprings is filled, we select the new population according the "$\alpha+\mu$" strategy which are commonly used in the evolution strategy community [12]. We select into the new population individuals from both of parent and offspring sets. Moreover, we use elitism strategy [12] as well, where we put the best individual into the new population in advance. As the termination criterion we used predetermined number of population exchanges.

It is clear, that GA contains many control parameters which affect its efficiency. It is needed to suitably set these parameters.

4 Numerical experiments

We implemented and tested the suggested GA to obtain the upper bound of RVAP. To verify the suggested method we used twenty benchmarks of evacuation plan design problem. These instances were created on the transportation network of Slovak Republic. We performed experiments on a personal computer equipped with Intel Pentium D with parameters 3 GHz CPU and 1 GB RAM. In the first series we used the special rounding heuristics only to obtain the upper bound of RVAP in BB method. In the tables this heuristic is denoted with symbol *RH* (Rounding Heuristics). In the second series we incorporate this heuristics into GA and GA was used to obtain the upper bound. In the tables GA is denoted with symbol *GA*. Since we used PMX scheme for crossover, we set the number of crossover points to value of two. When the mutation is carried out, only one change of genes are performed.

During second series of experiments, we set the crossover rate χ-and mutation rate μ to some different values. Beasley *et al*. Beasley *et al*. [1] recommends carrying out the crossover with probability about 80-95 % and the mutation with probability about 0.5-1 %. According to this, we successively set χ to values of 0.8, 0.85, 0.9, and 0.95 and μ to values of 0.005, 0.01, 0.015, and 0.02. Number of population (*NoP*) which served as termination criterion was successively set to values of 20, 40, and 60. Number of individuals (*NoI*) in each population

was successively set to values of 20, 40, and 60 as well. Each combination of these parameters (χ, μ, *NoP*, and *NoI*) values was used in the second series. The experiments were evaluated with *Data Precedence Analysis* (DPA) method [9]. The best parameters setting (PS) was $\chi = 0.85$, $\mu = 0.005$, $NoP = 20$, and $NoI = 20$. The values of χ and μ correspond to recommended values. But even with this best PS of GA we obtained worse results than with RH. The **Table 1** shows only benchmarks where the different evacuation time (*T*) was achieved for RH and GA. Only for the benchmark 11 and 15 were achieved better results when GA was used.

Benchmarks		02	03	04	05	06	07	09	11	13	15	16	19
T [min]	*RH*	278	69	129	61	95	150	92	128	124	457	220	61
	GA	284	70	131	65	97	160	94	127	125	455	223	62

Table 1 Result of RH and GA

Therefore, we decided to perform new experiments. We fixed χ and μ to values of 0.85 and 0.005 respectively and tried to explore efficiency of GA for wider range of the *NoP* and *NoI* parameters. We set successively *NoP* to values 10, 20, 50, and 100 and *NoI* to values 10, 20, 50, 100, and 150. We again solved benchmarks for every combination of *NoP* and *NoI*. Unambiguously, the best PS was for $NoP = 10$ and $NoI = 10$. These results invoked any suspicion, because the expectation was that the more the values of *NoP* and *NoI* are, the better results we should obtain. But the experiments affirmed the opposite.

So we decided to explore the efficiency for lesser values of *NoP* and *NoI*. We set successively *NoP* to values 1, 5, 10, and 15 and *NoI* to values 2, 5, 10, 15, and 20 and again solved the benchmarks for every PS combination. The PS, where $NoP = 1$ and $NoI = 2$, is unambiguously the best PS and almost the same results were achieved with this PS of GA as with RH. The second-best combinations of PS are shown in the **Table 2** with the best PS together.

Parameters setting of GA	Best	Second-best		
NoP	1	1	1	5
NoI	2	5	15	2

Table 2 Parameters setting of GA

The results show that the lesser the values of chosen parameters were, the better results we obtained. Based on these results we can conclude that it is not convenient to embed GA as a part of the branch and bound method. Although, GA is efficient at solving many problems, its power is based on the evolution process whereby the individuals, which represents problem solution, are bred. And this evolution process takes some portion of time (numbers of population exchanges). Moreover, population must contain sufficient number of individuals which depends on the size of problem. These factors influence the computational time which is needed for efficient performance of GA. But the methods which are used as a part of BB method have to work quickly, because the searching tree has big amount of nodes which have to be processed. For this reason we can not assign sufficient amount of resources (e.g. set *NoP* and *NoI* to sufficient high values) in order to obtain good results. On the contrary, for small values of *NoP* and *NoI* we are not able to reach good results and moreover the GA only slows the computational process. For this reason, lesser number of nodes in searching tree is processed for the same time with comparison with the case when GA is not used as a part of the BB method. In the case, when the searching process of the branch and bound method is prematurely terminated after the predetermined computational time expires, the searching process does not reach the branch (area in a searching tree) where the optimal solution would be found. Therefore we do not recommend embedding GA as a part of the branch and bound method.

5 Conclusion

In this paper we dealt with the evacuation plan design problem especially with the vehicle assignment problem which comprises its important part. Special iterative process is used to solve this problem where in each iteration the reduced problem is solved by the branch and bound method. We introduced the special rounding heuristics which is used to obtain the upper bound of the reduced problem in the original approach. We proposed to embed this heuristics within the genetic algorithm in order to obtain better results. The suggested method was experimentally verified by numerical experiments. The results show that it is not convenient to embed GA as a part of the branch and bound method because the power of GA is based on the evolution process which requires some resources (computation time). Quite the contrary, the methods used as a part of the branch and bound method have to work quickly.

Acknowledgements

This work was supported by the research grants VEGA 1/0296/12 "*Public Service Systems with Fair Access to Service*", APVV-0760-11 "*Designing of Fair Service Systems on Transportation Networks*" and by the institutional research grant of Faculty of Management Science an Informatics.

We would like to thank to "Centre of excellence in computer sciences and knowledge systems" (ITMS 26220120007) for built up infrastructure, which was used.

References

[1] Beasley, D., Bull, D. R., Martin, R. R.: An Overview of Genetic Algorithms: Part 1, Fundamentals. In *University Computing*. 1993. 15(2), pp. 58-69.

[2] FEMA. Evacuation Plans. http://www.fema.gov/plan/prepare/evacuation.shtm, 2012. [Online; accessed 11.5.2012].

[3] Janáček, J.: *Handling of a non-linear model of the evacuation plan design by IP-solver*. In *Proceedings of the 10th international Symposium on OPERATIONAL RESEARCH*. Nova Gorica, Slovenia : Slovenian Society Informatika, 2009. pp. 279-287.

[4] Janáček, J.: *Matematické programování.* 2nd revised edition. Žilina : Žilinská univerzita, 2003. 225 p.

[5] Janáček, J.: Tvorba evakuačních plánů pomocí na trhu dostupných IP-solverů. In *Úlohy diskrétní optimalizace v dopravní praxi*. Pardubice : Univerzita Pardubice, 2009. pp. 50-60.

[6] Janáček, J., Sibila, M.: *Optimal evacuation plan design with IP-solver*. In COMMUNICATIONS. ISSN 1335-4205, 2009, vol. 11, č. 3, pp. 29-35.

[7] Reeves, C. R.: Genetic Algorithms. In GENDREAU, M., POTVIN, J.-Y. (eds.) *Handbook of Metaheuristics*. New York: Springer, 2010. pp. 109-139.

[8] Toman, Ľ.: Acceleration of evacuation problem solving by branch and bound method with assistance of rapid excluding of branches. In *Mathematical methods in economics 2011: proceedings of the 29th international conference*. Praha: Professional Publishing, 2011. pp. 715-720.

[9] Toman, Ľ.: Data Precedence Analysis – Method for Evaluating of the Optimization Methods Parameters Settings. In *MICT 2012, 7th Winter School of Mathematics applied to ICT*. Žilina: Žilinská univerzita, 2012. [to be published].

[10] Toman, Ľ.: Min-Max Approach for Evacuation Problem Supported by BB-Search. In *Transcom : 9-th EUROPEAN CONFERENCE OF YOUNG RESEARCH AND SCIENCE WORKERS*. Slovak republic, Žilina : University of Žilina, 2011. pp. 215-218.

[11] Yusoff, M., Ariffin, J., Mohamed, A.: An Improved Discrete Particle Swarm Optimization in Evacuation Planning. In: *International Conference of Soft Computing and Pattern Recognition*. IEEE Computer Society, Malaysia. 2009. pp. 49-53.

[12] Zelinka, I. *et al*.: *Evoluční výpočetní techniky*. 1st edition. Praha: BEN – technická literatúra, 2009. 534 p.

Hankel max-min matrices and their applications

Hana Tomášková[1], Martin Gavalec[2]

Abstract. Hankel matrices are characterized as square matrices with constant entries along every skew-diagonal. Therefore, a Hankel matrix is determined by its entries in the first row and the last column. Special subtype is represented by circulant-Hankel matrices, with entries in every row shifted by one position to the left, in comparison with the previous row. A circulant-Hankel matrix is determined by entries in the first row. Properties of Hankel matrices and circulant-Hankel matrices in max-min algebra are studied in the paper. The question of finding the steady states of complex systems (eigenvectors of the transition matrix) is considered. Specific cases with the maximal matrix input value on different positions are completely described. Applications to real problems, e.g. the influence of viral advertising are presented.

Keywords: max-min algebra, Hankel matrices, spreading of information, marketing

JEL classification: C44
AMS classification: 90C15

1 Introduction

Many applications can naturally be solved by using operations maximum and minimum on a linearly ordered set. The so-called max-min algebras have been frequently studied by number of authors and various algorithms have been developed.

Eigenvectors of a max-min matrix characterize stable states of the corresponding discrete-events system. Investigation of the max-min eigenvectors of a given matrix is therefore of great practical importance. The eigenproblem in max-min algebra has been studied by many authors. Interesting results were found in describing the structure of the eigenspace, and algorithms for computing the maximal eigenvector of a given matrix were suggested, see e.g. [1], [2], [4], [8], [9], [10]. The structure of the eigenspace as a union of intervals of increasing eigenvectors was described in [3].

By max-min algebra we understand a triple $(\mathcal{B}, \oplus, \otimes)$, where $\mathcal{B}$ is a linearly ordered set, and $\oplus = \max$, $\otimes = \min$ are binary operations on $\mathcal{B}$. The notation $\mathcal{B}(n,n)$ $(\mathcal{B}(n))$ denotes the set of all square matrices (all vectors) of given dimension n over $\mathcal{B}$. Operations $\oplus$, $\otimes$ are extended to matrices and vectors in a formal way.

The eigenproblem for a given matrix $A \in \mathcal{B}(n,n)$ in max-min algebra consists of finding a vector $x \in \mathcal{B}(n)$ (eigenvector) such that the equation $A \otimes x = x$ holds true. By the eigenspace of a given matrix we mean the set of all its eigenvectors.

In this paper the structure of eigenvectors of Hankel matrices is studied. The advantage of the special matrices is that they enable more efficient solution of some specific problems. Hankel matrices are a special form of matrices, which are determined by the first row and the last column of matrix. The paper presents a detailed description of some possible types of eigenvectors of any given Hankel matrix.

[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

2 Application

As application of the Hankel matrices we present a simplified model of the information flow in viral marketing. By the source of disseminated information, the messages are divided into groups in which the degree of suggestibility can be monitored. For the sake of simplicity we only work with simplified linear model, which can be described by infinite Hankel matrix. In real application we are working with a finite segment of the infinite linear model, which means its reduction to a finite Hankel matrix.

The object of the model is spreading of information during an advertising campaign in mobile or internet environment, using a mechanism of the viral marketing. A classic environment suitable for application of viral marketing are social networks, such as Facebook, Twitter, Google +, etc.

In the model, transition matrix is considered with entries evaluating the intensity (the importance, etc.) of the links (follow-ups, coherences, etc.) between communication nodes. For any i, j the value a_{ij} is the intensity of the interconnection of the network node i with node j, and a_{ji} is the intensity of the interconnection of node j with node i. The values a_{ij} and a_{ji} must be equal, because they express the intensity of the interconnection, which is the same in both directions when using contemporary mobile technologies. Moreover, the intensity only depends on the source of the disseminated information, thus creating the pairs obtaining the particular information in the same time (with equal communication distance from the source).

Thus, in the linear model we get formal description by Hankel transition matrices which, due to their special form, represent spreading of information with constant intensity in pairs with equal communication distance from the source of information. The transition matrix A transforms the initial state vector x to a new state vector $A \otimes x$, where the state vector entries represent the impact degree of the marketing campaign on the communication nodes.

2.1 Model of suggestibility in viral advertising campaigns

Let us consider n interest groups (nodes in the model) such as they can be found in the social network Facebook. For each group we determine the degree of suggestibility in a viral advertising campaign. The rate can be calculated

$$\text{rate} = \frac{\text{number of advertising messages sent per week in group}}{\text{number of members}} .$$

In addition, we determine the level of mutual suggestibility communications between different groups

$$\text{suggestibility level} = \frac{\text{number of advertising messages sent between the two groups per week}}{\text{number of members of both groups}} .$$

Furthermore, we set the initial impact degree of the campaigne on each node in the model. The impact degree on each group lies within the interval $\langle 0; 1 \rangle$. Level 1 is assigned to a group (node) that is fully affected by the advertising campaign, while the group gets level 0, when the campaign impact on the group is neglectible.

The primal goal of the marketing campaign is to keep the present customers and avoid the unstable behaviour in the market. The stable states of the system correspond to eigenvectors of the transition matrix. For Hankel matrices the eigenproblem can be solved efficiently by considering only the entries in the first row and the last column of the matrix.

By the previous state of the art, it was only known that every constant vector with the value between the minimum and the maximum of the matrix entry values is a stable vector. The disadvantage is that constant eigenvectors can lead either to low penetration in the given marketing segment, or to loss of time and unnecessary costs in addressing the groups that are not advertising enough to hit other groups, or do not use sufficiently many links with other groups. The methods of extremal algebra enable us to describe further eigenvectors, thus helping us to find further (possible all) stable states of the system and increase the efficiency of the viral marketing campaign.

3 Eigenvectors of Hankel matrices

A square matrix is called Hankel matrix is a square matrix with constant skew-diagonals. Hence, Hankel matrix A is fully determined by its inputs in the first row and in the last column.

We shall use the notation $a(n-1) := a_{11}$, $a(n-2) := a_{12}, \dots, a(1) := a_{1n-1}$, $a(0) := a_{1n}$, $a(-1) := a_{2n}$, $\dots$, $a(2-n) := a_{n-1\,n}$, $a(1-n) := a_{nn}$.

$$\begin{pmatrix} a(n-1) & a(n-2) & \dots & a(1) & a(0) \\ a(n-2) & a(n-3) & \dots & a(0) & a(-1) \\ a(n-3) & a(n-4) & \dots & a(-1) & a(-2) \\ \vdots & \vdots & \dots & \vdots & \vdots \\ a(2) & a(1) & \dots & a(4-n) & a(3-n) \\ a(1) & a(0) & \dots & a(3-n) & a(2-n) \\ a(0) & a(-1) & \dots & a(2-n) & a(1-n) \end{pmatrix}$$

We shall use the notation $N^\star = \{1-n, 2-n, \dots, -1, 0, 1, \dots, n-2, n-1\}$. Further we define, for a Hankel matrix A given by the entries $\big(a(k),\, k \in N^\star\big)$ a strictly decreasing sequence $M(A) = (m_1, m_2, \dots)$ of length $s(A)$ by recursion

$$m_t = \begin{cases} \max\{\, a(k);\, k \in N^\star \,\} & \text{for } t = 1 \\ \max\{\, a(k) < m_{t-1};\, k \in N^\star \,\} & \text{for } t > 1 \end{cases}$$

Our research of Hankel matrices is aimed on complete description of all eigenvectors of a given Hankel matrix. The form of the eigenvectors strongly depends on positions of values $m_1, m_2, \dots$ in the entry vector $\big(a(k),\, k \in N^\star\big)$. Three interesting partial cases of our investigation are presented below.

Case 1

In this case we consider a matrix containing the maximum m_1 at one position only, with the remaining matrix entries equal to $m_2 = 0$. All possible eigenvectors are of the form indicated in the table below. In dependence on the position of the maximum entry, the entries of every eigenvector are divided into two groups. In every group the entries can have arbitrary values in the indicated interval, and, with one exception, the values must be symmetric within every entry group.

Position	Symmetry	Value interval	Entry group	Value interval	Entry group
$a(0)$	Yes	$\langle m_1; m_2 \rangle$	$\langle x_1; x_n \rangle$	-	-
$a(n-k); k \neq 1, n$	Yes	$\langle m_1; m_2 \rangle$	$\langle x_1; x_k \rangle$	$\{0\}$	$\langle x_{k+1}; x_n \rangle$
$a(n-1)$	No	$\langle m_1; m_2 \rangle$	x_1	$\{0\}$	$\langle x_2; x_n \rangle$
$a(1-k); k \neq 1, n$	Yes	$\langle m_1; m_2 \rangle$	$\langle x_{k+1}; x_n \rangle$	$\{0\}$	$\langle x_1; x_{k-1} \rangle$
$a(1-n)$	No	$\langle m_1; m_2 \rangle$	x_n	$\{0\}$	$\langle x_1; x_{n-1} \rangle$

Case 2

In this case we consider a matrix containing the maximum m_1 on two positions, with the remaining matrix entries equal to $m_2 = 0$. The symmetry must now occur simultaneously on two (sometimes only one) entry intervals. If the entry symmetry intervals do not cover the entire vector, the remaining area can only contain zero entries.

Positions	Value interval	Entry symmetry interval
$a(0), a(n-k); k \neq 1, n$	$\langle m_1; m_2 \rangle$	$\langle x_1; x_n \rangle \wedge \langle x_1; x_k \rangle$
$a(0), a(n-1)$	$\langle m_1; m_2 \rangle$	$\langle x_1; x_n \rangle$
$a(n-k), a(n-l); k < l < n$	$\langle m_1; m_2 \rangle$	$\langle x_1; x_k \rangle \wedge \langle x_1; x_l \rangle$
$a(0), a(1-k)$	$\langle m_1; m_2 \rangle$	$\langle x_1; x_n \rangle \wedge \langle x_k; x_n \rangle$
$a(0), a(1-n)$	$\langle m_1; m_2 \rangle$	$\langle x_1; x_n \rangle$
$a(1-k), a(1-l); k < l < n$	$\langle m_1; m_2 \rangle$	$\langle x_k; x_n \rangle \wedge \langle x_l; x_n \rangle$

Case 3

In this case we consider a matrix containing two different maximums m_1, m_2, while the remaining entries are equal to $m_3 = 0$. If the vector field does not cover the entire vector, the remaining area contains only zeros. Vector interval corresponding to the maximum m_1 must be symmetric in values and in positions. If the vector interval corresponding to the second maximum m_2 must be symmetric in positions, but two cases of values greater then m_2 must be considered.

1. Position of m_2 is smaller then position of m_1, then the values on symmetric position should be contained in interval $\langle m_1; m_2 \rangle$.
2. Position of m_2 is greater then position of m_1, then the values on symmetric position must be equal to m_2.

For the inputs on interval $\langle m_2; m_3 \rangle$ the symmetry is requuired on both vector intervals.

4 Comparison with other special matrices

The circulant, Toeplitz, Hankel and circulant-Hankel matrices are often studied as important special matrices in extremal algebras. The matrices have interesting applications, and on the other hand, many algorithms have lower computational complexity in these special cases.

A circulant matrix has a special form given by the entries in the first row. Next rows contain the same values as the first one, but the values are shifted one position to the right. The characteristic digraph of the matrix is symmetric with respect to cyclic permutation of nodes.

Toeplitz matrices are generalization of circulant matrices. A Toeplitz matrix contains constant values on every line parallel with the main diagonal. Hence, the matrix is given by its first row and its first column. Toeplitz matrix can be interpreted as a finite part of an infinite circulant matrix.

Hankel matrices are in some sense dual to Toeplitz matrices, namely they contain constant values on every line parallel with the skew diagonal. Dualy to Toeplitz matrices, all entries in a Hankel matrix are determined by the entries in the first row and in the last column.

Finally, circulant-Hankel matrices represent a special case of Hankel matrices. A circulant-Hankel matrix is given by entries in the first row, and the next rows contain the same values shifted to one position to the left.

There exist also further connection between these special matrix types. In particular, the second power of Hankel matrix is a Toeplitz matrix and third power is again a Hankel matrix. The same connection holds between circulant-Hankel matrices and circulant matrices. As a consequence, an eigenvector of a Hankel matrix is also eigenvector of some Topelitz matrix and an eigenvector of a circulant-Hankel matrix is also eigenvector of some circulant matrix. These connections are illustrated by examples in this section.

Example 1. Let Hankel matrix A with one maximal value on the input $a(0)$ be given.

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 10 \\ 0 & 0 & 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 10 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 & 0 & 0 \\ 10 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Eigenvectors corresponding to this matrix are all symmetric vectors containing arbitrary values from interval $\langle 0, 10\rangle$, i.e. vectors of the form $(x_1, x_2, x_3, x_3, x_2, x_1)$.

The second power of Hankel matrix A is a Toeplitz (and in this case also circulant) matrix

$$A^2 = \begin{pmatrix} 10 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 & 0 \\ 0 & 0 & 0 & 10 & 0 & 0 \\ 0 & 0 & 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 0 & 0 & 10 \end{pmatrix}$$

Eigenvectors of this matrix are all vectors containing arbitrary values from the interval $\langle 0, 10\rangle$, as it was proved in [5]. Hence, in this example all eigenvectors of the given Hankel matrix A are also eigenvectors of the circulant matrix A^2, but the converse implication does not hold. Some of the eigenvectors of circulant matrix A^2 (the non-symmetric ones) are not eigenvectors of Hankel matrix A.

Example 2. Let Hankel matrix B with maximum on two positions $a(0), a(3)$ be given.

$$B = \begin{pmatrix} 0 & 0 & 10 & 0 & 0 & 10 \\ 0 & 10 & 0 & 0 & 10 & 0 \\ 10 & 0 & 0 & 10 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 & 0 & 0 \\ 10 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Eigenspace corresponding to matrix B consists of all vectors symmetric on entry interval $[x_1, x_3]$ and at the same time symmetric on the whole entry interval $[x_1, x_6]$. Under these symetry conditions the eigenvectors can contain arbitrary entry values from interval $\langle 0, 10\rangle$. Hence, every eigenvector of B is of the form $(x_1, x_2, x_1, x_1, x_2, x_1)$ with arbitrary values $x_1, x_2 \in \langle 0, 10\rangle$.

The second power of Hankel matrix B is a Toeplitz (and, similarly as in the previous example, also circulant) matrix

$$B^2 = \begin{pmatrix} 10 & 0 & 0 & 10 & 0 & 0 \\ 0 & 10 & 0 & 0 & 10 & 0 \\ 0 & 0 & 10 & 0 & 0 & 10 \\ 10 & 0 & 0 & 10 & 0 & 0 \\ 0 & 10 & 0 & 0 & 10 & 0 \\ 0 & 0 & 10 & 0 & 0 & 10 \end{pmatrix}$$

Eigenspace of circulant matrix B^2 consists of all eigenvectors containing values form interval $\langle 0, 10\rangle$ which must be repeated every third position, because the greatest common divisor of maximum position and size of matrix is $\gcd(3, 6) = 3$, see [5]. Hence the eigenvectors of B^2 are all vectors of the form $(x_1, x_2, x_3, x_1, x_2, x_3)$ with arbitrary values $x_1, x_2, x_3 \in \langle 0, 10\rangle$.

Similarly as in the previous example, all eigenvectors of the given Hankel matrix B are also eigenvectors of the circulant matrix B^2, but the converse implication does not hold.

Example 3. Let C be a given Hankel matrix with maximum on two positions $a(0), a(2)$.

$$C = \begin{pmatrix} 0 & 0 & 0 & 10 & 0 & 10 \\ 0 & 0 & 10 & 0 & 10 & 0 \\ 0 & 10 & 0 & 10 & 0 & 0 \\ 10 & 0 & 10 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 & 0 & 0 \\ 10 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Eigenspace corresponding to matrix C only contains constant vectors with value from interval $\langle 0, 10\rangle$.

Let us remark that in this example the second power of the given matrix is a Toeplitz matrix which in not circulant.

$$C^2 = \begin{pmatrix} 10 & 0 & 10 & 0 & 0 & 0 \\ 0 & 10 & 0 & 10 & 0 & 0 \\ 10 & 0 & 10 & 0 & 10 & 0 \\ 0 & 10 & 0 & 10 & 0 & 10 \\ 0 & 0 & 10 & 0 & 10 & 0 \\ 0 & 0 & 0 & 10 & 0 & 10 \end{pmatrix}$$

Eigenspace of Toeplitz matrix C^2 consists of all eigenvectors containing values form interval $\langle 0, 10\rangle$ which must be repeated every second position, because the greatest common divisor of maximum position in the first row and the size of matrix is $\gcd(4, 6) = 2$, see [6]. Hence the eigenvectors of C^2 are all vectors of the form $(x_1, x_2, x_1, x_2, x_1, x_2)$ with arbitrary values $x_1, x_2 \in \langle 0, 10\rangle$.

ACKNOWLEDGEMENT

The support of INKOV project # CZ.1.07/2.3.00/20.0001 financed from EU and Czech Republic funds, and Czech Science Foundation project # 402/09/0405 is kindly acknowledged.

(Received July 9, 2012.)

References

[1] K. Cechlárová: Eigenvectors in bottleneck algebra. *Lin. Algebra Appl.* **175** (1992), 63–73.

[2] R.A. Cuninghame-Green: *Minimax Algebra and Application.* In: Advances in Imaging and Electron Physics 90, (P. W. Hawkes, Ed.). Academic Press, New York, 1995.

[3] M. Gavalec: Monotone eigenspace structure in max-min algebra. *Lin. Algebra Appl.* **345** (2002), 149–167.

[4] M. Gavalec, J. Plavka: Eigenproblem in extremal algebras. *Proc. of the 9th Int. Symp. Oper. Res. '07,* Nova Gorica, Slovenia, 2007.

[5] M. Gavalec, H. Tomášková: Eigenspace of a circulant max-min matrix. *Kybernetika* **46**(3), 397–404.

[6] M. Gavalec, H. Tomášková: Eigenproblem for circulant and Hankel matrices in extremal algebra. *Proc. of SIAM '12*, Valencie, Spain, 2012.

[7] R.M. Gray: *Toeplitz and Circulant Matrices*, NOW Publishers, Delft, 2006.

[8] J. Plavka: Eigenproblem for circulant matrices in max-algebra. *Optimization* **50** (2001), 477–483.

[9] J. Plavka: l-parametric eigenproblem in max algebra. *Discrete Applied Mathematics* **150** (2005), 16–28.

[10] K. Zimmermann: *Extremal Algebra* (in Czech). Ekon. ústav ČSAV Praha, 1976.

Comparison of two different approaches to stock portfolio analysis

Filip Tošenovský[1]

Abstract. The article presents a portfolio analysis of selected shares traded on Prague Stock Exchange, using two approaches. One approach is represented by the standard procedure which defines a portfolio on the basis of optimization of the trade-off between the first and second moments of the considered share prices, with the moments being estimated from the available data by their common sample counterparts. The second approach to the analysis is based on the estimates capturing the time dynamics of the first and second moments of the share prices that are typically generated by nonstationary processes. To capture the dynamics, time series models are used. The two procedures are compared as to which of them provides an expected portfolio return that is closer to the true return.

Keywords: stock portfolio, moment estimates, time series.

JEL Classification: C13, C61
AMS Classification: 91G10, 91G70

1 Introduction

Stock market has long been in the spotlight of many investors, as it enables flexible changes in the financial structure of private entities, and provides room for investments. Investments especially boost a lot of interest in part due to theories on stock portfolio optimization. The theories try to reduce the risk resulting from the fluctuation of stock prices. There is more than one approach to achieving such objective. The procedures often use stochastic calculus, which is the approach we adopt in this article as well.

Creating a portfolio falls in the category of optimization, so that the selection of shares forming the portfolio must necessarily follow certain criteria based on which the portfolio is optimized. One should therefore pay attention to these criteria for their irrational selection may provide the portfolio owner with dissatisfactory results despite the fact that the rest of the portfolio formation was taken care of by exact mathematics. If we focus on the classical way of creating a portfolio, as proposed by Markowitz, the optimization criteria used are the unconditional expected value and variance of the portfolio yield. However, in reality their estimates must be used, which means to estimate unconditional expected values of the stock yields and covariances of these yields.

The commonly used estimates are calculated as in the case of estimation of stationary process moments. The question then is what is actually calculated by the common estimates, especially when stock prices are typically generated by nonstochastic processes. Apart from the way these estimates are obtained, their inertia is assumed as well for the time period during which the portfolio to be optimized is held by its owner. These are the reasons why we digress to other portfolio criteria in this paper – the predicted conditional expected values and variances of the stock yields with the predictions carried out using time series models. The alternative approach tries to overcome the inertia assumption, taking into account the latest dynamics of the market. We shall use both the standard and alternative approach for the portfolio optimization, and compare the yields of such portfolios.

In this paper, we shall draw the analysis on the classical Markowitz's approach to portfolio optimization. This approach aims to minimize the estimate of the portfolio unconditional variance $var[f(w_1, w_2, \dots, w_n)] = \sum_{i,j=1}^{n} w_i w_j s_{ij}$ with respect to the variables w_i's. Here, $f(w_1, w_2, \dots, w_n) = \sum_{i=1}^{n} w_i x_i$ represents the portfolio yield, x_i is a yield of the *i*-th stock included in the portfolio, w_i is the weight of the *i*-th stock in the portfolio and s_{ij} denotes an estimate of the covariance between the yields of the *i*-th and *j*-th stock. The estimate of the unconditional expected return of the portfolio equals $\sum_{i=1}^{n} w_i \bar{x}_i$, $\bar{x}_i$ being an estimate of the unconditional average yield of the *i*-th stock. The extreme of the function *var f* is sought in the set *M* defined by (in)equalities: $\sum_i w_i = 1$, $w_i \geq 0$ for $i = 1, 2, \dots, n$ and $\sum_{i=1}^{n} w_i \bar{x}_i \geq k$, where *k* is defined arbitrarily.

[1] Silesian University in Opava/School of Business Administration in Karvina, Department of Mathematical Methods in Economics, Univerzitní náměstí 1934/3, 733 40 Karviná, the Czech Republic, tosenovskyfilip@opf.slu.cz.

Generally, the problem with estimates in the portfolio optimization is as follows: if we had perfect estimates of the moments of the future stock yields, the optimization based on Markowitz would mean that we could rely on having a *decent realized* investment yield not far from the expected yield of the portfolio. This is assured by setting up the portfolio so that its riskiness/variance is minimized while a certain predefined decent level of expected yield is assured. However, the estimates will almost surely not be perfect. Thus, we minimize $\sum_{i,j=1}^{n} w_i w_j s_{ij}$ instead of $\sum_{i,j=1}^{n} w_i w_j \sigma_{ij}$ on the set *M* differing from the original set *N* by the last inequality: $\sum_{i=1}^{n} w_i \bar{x}_i \geq k$ instead of $\sum_{i=1}^{n} w_i E(x_i) \geq k$, the latter defining together with the (in)equalities $\sum_i w_i = 1$, $w_i \geq 0$ the set *N*. This means that instead of the correct vector of weights $\boldsymbol{w}^{opt}$ we get a vector $\breve{\boldsymbol{w}}^{opt}$. Of course, it is very likely that $\breve{\boldsymbol{w}}^{opt}$ will not optimize $\sum_{i,j=1}^{n} w_i w_j \sigma_{ij}$ in *N* - if $\breve{\boldsymbol{w}}^{opt}$ belongs to *N,* that is. It may happen that $\breve{\boldsymbol{w}}^{opt}$ does not belong to *N* at all because $\sum_{i,j=1}^{n} \breve{w}_i^{opt} E(x_i) < k$. The better the moment estimates, the more the portfolio should be under control in the sense that its *realized* yield should be closer to its expected yield.

In light of what was just said, we will now present results of an empirical study that compares the *two approaches* mentioned in the introduction. To do so, we shall draw on the data from the Prague Stock Exchange.

2 Classical approach

For the purpose of our analysis, we shall now employ stocks of five companies with the highest market value listed in the SPAD trade system. These are, as implied by the Prague Stock Exchange data, Central European Media Enterprises (CEME), ČEZ, Erste Bank, Telefónica and NWR. We will use the closing prices of the stock from the start of 2012 to April of that year. The data are at www.penize.cz. Table 1 shows some of the oldest and latest prices of the stock for the given period.

Date	CEME	CEZ	ERSTE	NWR	TELEFÓNICA
27.04.12	146.06	758.0	426.0	126.53	379.0
26.04.12	139.50	759.9	423.3	126.10	378.0
25.04.12	134.10	764.9	435.6	125.20	378.1
…	…	…	…	…	…
04.01.12	134.60	805.0	353.7	142.95	391.5
03.01.12	138.00	800.9	366.5	140.80	385.5
02.01.12	133.64	791.0	359.0	139.29	382.5

Table 1 The latest and oldest closing prices of the selected stock (in Czech crowns)

We shall assume the investor always wants to set up the portfolio for a week. More precisely, the portfolio is created on Friday, and kept until next Friday when its realized profit/loss is confronted. Then the investor sets up a new weekly portfolio, taking into account the new data on the market development. Thus a series of portfolios whose composition differs in time is created. To optimize the portfolios, we first calculate the standard estimates of the first and second moments of the stock yields, i.e. sample averages and sample variances of the five-day yields of each stock and sample covariances between the yields, using historical data. To get reasonable estimates, we start with the time series of the oldest fifty values. Thus, the first and the oldest portfolio is optimized on the estimates calculated from the data from the start of 2012 to March 9. The second portfolio uses estimates calculated from the data from the start of 2012 to March 16, and so on. Table 2 shows sample averages of the yields used for seven subsequently optimized weekly portfolios.

Week	CEME	CEZ	ERSTE	NWR	TELEF.
1.	0.003457	0.003061	0.035187	0.006552	0.003778
2.	0.007675	0.002128	0.029835	0.001891	0.001979
3.	0.005171	0.002481	0.028431	-0.00546	0.001668
4.	0.003155	0.001549	0.02236	-0.00266	0.001223
5.	0.007269	0.00062	0.014641	-0.00529	-0.00052
6.	0.003431	-0.00032	0.011721	-0.00596	-0.00201
7.	0.003148	-0.00167	0.0114	-0.00578	-0.00101

Table 2 Sample averages of five-day yields to be used successively for seven weekly portfolios

Table 3 describes sample variances of the yields for different weeks, table 4 shows the covariance structure of the yields for the first constructed portfolio (covariances for other weeks were obtained in a similar fashion).

Week	CEME	CEZ	ERSTE	NWR	TELEF.
1.	0.007939	0.000722	0.007244	0.003090	0.000811
2.	0.007477	0.000663	0.006820	0.003027	0.000763
3.	0.006895	0.000610	0.006244	0.003327	0.000699
4.	0.006410	0.000570	0.006165	0.003170	0.000643
5.	0.006200	0.000542	0.006434	0.003011	0.000654
6.	0.006095	0.000527	0.006300	0.002896	0.000670
7.	0.005690	0.000526	0.005899	0.002722	0.000643

Table 3 Sample variances of five-day yields to be used successively for seven weekly portfolios

	CEME	CEZ	ERSTE	NWR	TELEF.
CEME	0.007939	0.001483	0.004958	0.002092	-0.00065
CEZ	0.001483	0.000722	0.000449	0.000617	0.000187
ERSTE	0.004958	0.000449	0.007244	0.001194	-0.00038
NWR	0.002092	0.000617	0.001194	0.00309	0.000162
TELEF.	-0.00065	0.000187	-0.00038	0.000162	0.000811

Table 4 Sample covariances of five-day yields to be used for the first portfolio

Having the estimates, we performed the optimization in the Markowitz's sense, and we arrived at the results described in table 5. Expected return at the optimum for each week is on the far left of the table, realized return is the return the investor actually registered during the five-day period of holding the newest portfolio. Absolute difference of the two returns is contained in the third column of the table. The last five columns portray the portfolio weights at the optimum. As one can see, some of the differences in returns are quite severe.

Expected Return	Realized Return	Abs. Difference	w(1)	w(2)	w(3)	w(4)	w(5)
0.003	-0.0142	0.0174	0.000	0.495	0.045	0.033	0.428
0.010	0.0004	0.0096	0.000	0.182	0.412	0.000	0.406
0.010	-0.0284	0.0384	0.000	0.170	0.440	0.000	0.390
0.010	-0.0797	0.0897	0.000	0.000	0.594	0.000	0.406
0.005	0.0110	0.0060	0.048	0.341	0.484	0.000	0.127
0.005	0.0181	0.0131	0.000	0.416	0.584	0.000	0.000
0.005	0.0310	0.0260	0.000	0.000	0.615	0.000	0.385

Table 5 Portfolio optimization based on standard moment estimates, and the true development of the market

3 Alternative approach

We shall now employ the alternative procedure in the portfolio analysis. We will not assume for each weekly upgrade of the portfolio that the expected return of the stock yield and its variance do not undergo any dynamic change, as was assumed in the classical procedure. We shall assume the contrary, and we will describe the dynamics with ARIMA model family [1]. Each model used depicts the time development of the five-day yields of the given stock. The model was successively updated with the arrival of new data, however we updated it only every other week because one week of new data was a too short period to lead to a perceptible modification of the model. We found each model on the basis of Akaike information criterion [2], using autocorrelation and partial autocorrelation functions as the first hint for the selection of the model. There are usually several candidates for a good model. In most cases, an autoregressive model proved to be the best by the information criterion. In several cases, especially in the case of Erste Bank stock, moving average model proved to be better by the criterion, however the model was noninvertible with one or more of the roots of the moving average lag polyno-

mial equal to one. Since such model would be difficult to use for forecasts, we selected also in this case an autoregressive model which was one of the best by the information criterion, although not the very best. The autoregressive models were applied to the differenced series of the five-day stock yields. Resulting models containing only statistically significant coefficients are as follows:

First and Second Week:
CEME: $\Delta z_t = -0.609\Delta z_{t-4} - 0.408\Delta z_{t-8} + 0.317\Delta z_{t-9} + \varepsilon_t$
ČEZ: $\Delta z_t = -0.251\Delta z_{t-1} - 0.471\Delta z_{t-4} - 0.354\Delta z_{t-5} + \varepsilon_t$
ERSTE: $\Delta z_t = -0.445\Delta z_{t-4} + \varepsilon_t$
NWR: $\Delta z_t = -0.46\Delta z_{t-4} - 0.471\Delta z_{t-8} + 0.41\Delta z_{t-12} + \varepsilon_t$
TEL : $\Delta z_t = -0.245\Delta z_{t-1} - 0.394\Delta z_{t-4} + \varepsilon_t$

Third and Fourth Week:
CEME: $\Delta z_t = -0.627\Delta z_{t-4} - 0.338\Delta z_{t-8} + \varepsilon_t$
ČEZ: $\Delta z_t = -0.23\Delta z_{t-1} - 0.75\Delta z_{t-4} + 0.34\Delta z_{t-5} + 0.24\Delta z_{t-7} - 0.35\Delta z_{t-8} + \varepsilon_t$
ERSTE: $\Delta z_t = -0.516\Delta z_{t-4} - 0.286\Delta z_{t-8} + \varepsilon_t$
NWR: $\Delta z_t = -0.462\Delta z_{t-4} - 0.467\Delta z_{t-8} - 0.4\Delta z_{t-12} + \varepsilon_t$
TEL: $\Delta z_t = -0.29\Delta z_{t-1} - 0.61\Delta z_{t-4} - 0.47\Delta z_{t-8} - 0.3\Delta z_{t-11} - 0.54\Delta z_{t-12} + \varepsilon_t$

Fifth and Sixth Week:
CEME: $\Delta z_t = -0.623\Delta z_{t-4} - 0.326\Delta z_{t-8} + 0.214\Delta z_{t-9} + \varepsilon_t$
ČEZ: $\Delta z_t = -0.693\Delta z_{t-4} + 0.264\Delta z_{t-7} - 0.277\Delta z_{t-8} + \varepsilon_t$
ERSTE: $\Delta z_t = -0.522\Delta z_{t-4} - 0.308\Delta z_{t-8} + \varepsilon_t$
NWR: $\Delta z_t = -0.49\Delta z_{t-4} - 0.506\Delta z_{t-8} - 0.36\Delta z_{t-12} + \varepsilon_t$
TEL: $\Delta z_t = -0.29\Delta z_{t-1} - 0.61\Delta z_{t-4} - 0.46\Delta z_{t-8} - 0.31\Delta z_{t-11} - 0.55\Delta z_{t-12} + \varepsilon_t$

Seventh Week:
CEME: $\Delta z_t = -0.638\Delta z_{t-4} - 0.319\Delta z_{t-8} + 0.21\Delta z_{t-9} + \varepsilon_t$
ČEZ: $\Delta z_t = -0.74\Delta z_{t-4} + 0.24\Delta z_{t-7} - 0.319\Delta z_{t-8} + \varepsilon_t$
ERSTE: $\Delta z_t = -0.527\Delta z_{t-4} - 0.346\Delta z_{t-8} + \varepsilon_t$
NWR: $\Delta z_t = -0.4976\Delta z_{t-4} - 0.353\Delta z_{t-8} + \varepsilon_t$
TEL: $\Delta z_t = -0.25\Delta z_{t-1} - 0.63\Delta z_{t-4} - 0.5\Delta z_{t-8} - 0.29\Delta z_{t-11} - 0.54\Delta z_{t-12} + \varepsilon_t$

These models were used first to obtain predictions of the conditional expected values of the five-days-ahead yields of each stock. Assuming independence of ε_t's across time, the expected value of z_t conditional on all the past available information z_{t-1}, z_{t-2},..., z_1 equals in case of a zero-mean stationary process $\Delta z_t = \varphi_1\Delta z_{t-1} + \varphi_2\Delta z_{t-2} + \cdots + \varphi_p\Delta z_{t-p} + \varepsilon_t$ to

$$\begin{aligned} E(z_t|z_{t-1}, z_{t-2}, \dots, z_1) &= z_{t-1} + \varphi_1\Delta z_{t-1} + \varphi_2\Delta z_{t-2} + \cdots + \varphi_p\Delta z_{t-p} + E(\varepsilon_t|z_{t-1}, z_{t-2}, \dots, z_1) \\ &= z_{t-1} + \varphi_1\Delta z_{t-1} + \varphi_2\Delta z_{t-2} + \cdots + \varphi_p\Delta z_{t-p} + E(\varepsilon_t|\varepsilon_{t-1}, \varepsilon_{t-2}, \dots) \\ &= z_{t-1} + \varphi_1\Delta z_{t-1} + \varphi_2\Delta z_{t-2} + \cdots + \varphi_p\Delta z_{t-p} + E(\varepsilon_t) \end{aligned}$$

or

$$E(z_t|z_{t-1}, z_{t-2}, \dots, z_1) = z_{t-1} + \varphi_1\Delta z_{t-1} + \varphi_2\Delta z_{t-2} + \cdots + \varphi_p\Delta z_{t-p}. \quad (1)$$

This is due to linearity of the expected value operator. Equation (1) and the law of iterated predictions [3] were used for the forecast of the conditional expected values of the five-days-ahead yields. Further, the models found were used to obtain estimates of the residuals $\hat{\varepsilon}_t$. The simple average of the second powers of these residuals was then employed to obtain the estimate of the variance $var(\varepsilon_t)$. Finally, using this estimate and the estimates of φ_i's, five-day-ahead conditional variances of the stock yields were calculated, using the usual operations applied to time series models. This means that if the model is $y_t = \varphi y_{t-4} + \varepsilon_t$, for instance, and one wants to calculate $var(y_{t+2}|y_t, y_{t-1}, \dots, y_1)$, then since $y_{t+2} = \varphi^2 y_t + \varphi\varepsilon_{t+1} + \varepsilon_{t+2}$ by recursive substitution, the conditional variance equals $var(\varepsilon_t)\cdot(1+\varphi^2) \cong \widehat{var}(\varepsilon_t)\cdot(1+\hat{\varphi}^2)$, as ε_t's are independent. The same logic was used in calculating $var(y_{t+5}|y_t, y_{t-1}, \dots, y_1)$. Since portfolio analysis also involves estimates of covariances, one might think of their dynamic description as well. A multivariate model might then be more appropriate for our problem, however the correct selection of such model is more involved and less certain. Therefore we adopted the standard estimates of the covariances here. Our alternative procedure is thus a hybrid one with

some moments being estimated in a nonstandard way, while others being estimated in the standard way. The forecasts of conditional expected values and variances are in table 6 and 7.

Week	CEME	CEZ	ERSTE	NWR	TELEF.
1.	-0.022	0.001	-0.007	-0.017	-0.015
2.	0.051	-0.010	0.012	-0.024	-0.015
3.	-0.004	0.010	-0.009	-0.033	-0.009
4.	-0.007	-0.005	-0.026	-0.042	-0.012
5.	-0.016	-0.010	-0.067	-0.037	-0.019
6.	0.017	-0.015	-0.022	0.016	0.005
7.	-0.012	-0.016	0.013	-0.009	-0.012

Table 6 Forecasts of five-day-ahead conditional expected yields of the selected stock for different weeks

Week	CEME	CEZ	ERSTE	NWR	TELEF.
1.	0.002510773	0.000337216	0.002696325	0.001061261	0.000562437
2.	0.002510773	0.000337216	0.002696325	0.001061261	0.000562437
3.	0.002667569	0.000305385	0.002096976	0.000915414	0.000360972
4.	0.002667569	0.000305385	0.002096976	0.000915414	0.000360972
5.	0.002206435	0.000260801	0.001856409	0.000926423	0.000321658
6.	0.002206435	0.000260801	0.001856409	0.000926423	0.000321658
7.	0.001955696	0.000288055	0.001797117	0.000946821	0.000280025

Table 7 Forecasts of five-day conditional variances of the yields for different weeks

Using the forecasts of conditional expected values and variances of the stock yields for a week ahead, the portfolio for each week was again optimized and the expected and actual return provided by the portfolio were compared. The results in decimal form are contained in table 8 which is an analogy to table 5.

Expected Return	Realized Return	Abs. Difference	w(1)	w(2)	w(3)	w(4)	w(5)
0.001	-0.0076	0.0086	0.000	0.975	0.000	0.000	0.025
0.010	-0.0105	0.0205	0.380	0.000	0.000	0.000	0.620
0.010	-0.0130	0.0225	0.000	1.000	0.000	0.000	0.000
-0.005	-0.0232	0.0187	0.000	1.000	0.000	0.000	0.000
-0.010	-0.0096	0.0003	0.000	1.000	0.000	0.000	0.000
0.007	0.0088	0.0019	0.176	0.000	0.000	0.000	0.824
0.005	0.0349	0.0299	0.000	0.000	0.687	0.000	0.313

Table 8 Portfolio optimization based on alternative moment estimates, and the true market progress

Let us note here that in reality a rational investor would not use the alternative approach in the fourth and fifth week because of the negative expected returns forecast. Since all of these returns are negative, as can be seen in table 6, there is obviously no way how to optimize the portfolio and not experience a loss. We included the two weeks in our analysis for the sole purpose of finding out how the estimates behave mathematically, as compared to the true returns in the fourth and fifth week.

Finally, Figure 1 compares the absolute differences between expectations and reality, recorded both in the case of the classical estimates and the alternative estimates. It can be seen that the alternative estimates perform strikingly better than the classical ones in the sense that in their case the absolute differences are smaller, enabling the investor to have a greater control over their portfolios. The better performance of the alternative estimates suggest that they are closer to the real expected value and variance of the portfolio held by the investor.

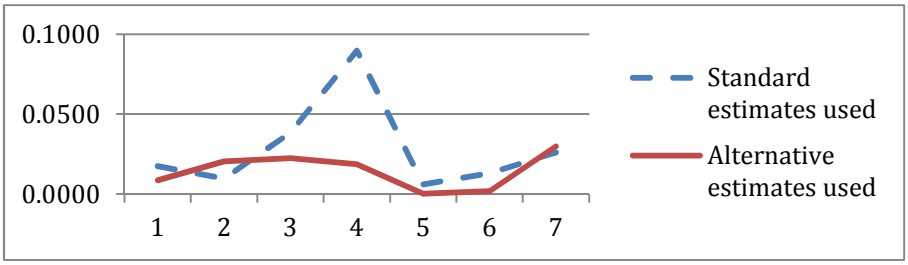


Figure 1 Absolute differences between expected returns and true returns for the period of seven weeks

Stock investments are accompanied by transaction costs which consist of the amount paid to the broker who carries out the trade, and the spread, i.e. the difference between the price at which the stock is bought, and the price at which the stock is sold. The spread is accounted for in the portfolio yields described in table 5 and table 8, so let us focus on the commission paid to the broker. The way the commission is calculated may vary among banks, but a common practice is to take it as a fixed percentage of the total amount invested in stock, regardless of whether the stock is bought or sold. Let S denote the initial amount of money invested in stock, and let $100p$ express the percentage put away for the commission, $1 > p > 0$. If the investor updates the portfolio once a week, the portfolio bought in the first week is valued at $S/(1+p)$. Depending on the development of the market, the value then changes to $[S/(1+p)] \cdot (1+i_1)$ where i_1 denotes the weekly yield of the portfolio. Updating the portfolio for the following week now means selling the current portfolio, i.e. obtaining the amount

$$[S/(1+p)] \cdot (1+i_1) - p[S/(1+p)] \cdot (1+i_1) = (1-p) \cdot [S/(1+p)] \cdot (1+i_1), \tag{2}$$

where the substraction occurs due to the transaction costs, and then reinvesting this amount to obtain a new portfolio of the desired composition, valued due to the transaction costs again at

$$\left(\frac{1-p}{1+p}\right) \cdot \left(\frac{S}{1+p}\right) \cdot (1+i_1). \tag{3}$$

Continuing this way, it is easy to see that after n weeks the investor registers in cash the amount

$$S \cdot \left(\frac{1-p}{1+p}\right)^n \cdot \prod_{k=1}^{n}(1+i_k), \tag{4}$$

where i_k denotes the portfolio weekly yield in the kth week of the investment. It is obvious from (4) that transaction costs play a role when one is concerned about the performance of the portfolio, but has no effect on the conclusion which of the *two portfolio strategies* we described in this article performs better, as the latter statement depends solely on the term $\prod_{k=1}^{n}(1+i_k)$. Using the real returns from table 5 and table 8, the term $\prod_{k=1}^{n}(1+i_k)$ equals 0.936 and 0.979, respectively, confirming that the alternative approach gives a better result.

4 Conclusion

We presented an analysis of two different approaches to optimizing a stock portfolio that would be always held for a period of five days. The optimization was performed as suggested by Markowitz. One of the approaches was the classical one, calculating the commonly used sample averages and covariances of five-day stock yields from the historical data and using them for the optimization. The alternative approach modelled the historical data with ARIMA equations, and calculated five-day forecasts of conditional averages and variances of the analyzed stock yields. The remaining second moments necessary for the optimization were estimated as in the classical case, i.e. using historical stock yields without using more involved ARIMA equations. The two approaches were compared as to which of them provides an expected return that is closer to the trully realized return. It turns out the alternative approach, benefiting from the description of the actual dynamics of the market, leads to an optimized portfolio with a more realistic expected return estimate. Investor holding such a portfolio has a greater control over the investment, and can rely to a greater extent on experiencing a true profit that is closer to the expected profit.

References

[1] Box, G. E. P., Jenkins, G. M., and Reinsel G.C.: *Time Series Analysis*. John Wiley & Sons, 2008.
[2] Brockwell P.J., and Davis, R.A.: *Time Series: Theory and Methods*. Springer-Verlag New York, 1991.
[3] Hamilton, J.D.: *Time Series Analysis*, Princeton University Press, 1994.

Multivariate time-series model of GDPs of the Czech Republic and its major economic partners

Filip Tošenovský[1], Elena Mielcová[2]

Abstract. This article applies time series theory to analyses of mutual relationships among gross domestic products of the Czech Republic and the countries serving as its major European trading partners. The article uses for this purpose multivariate time series models and data from the period of 1995-2011, as well as standard procedures to determine whether the model is stationary or not, and in the case of nonstationarity the nature of such instability. After finding the model, the time series model is tested for individual coefficient restrictions, and impulse-response dynamics are considered for each state as well.

Keywords: GDP, stationarity, multivariate model, impulse-response dynamic.

JEL Classification: C320
AMS Classification: 62P20

1 Introduction

The main objective of this article is to compare the development of the Czech economy with the economic development of several countries in the European Union in the period of 1995-2011. Economic development can be measured with different characteristics; this analysis is based on comparison of gross domestic products (GDP) of selected countries. In general, as stated in [4]: "*Gross domestic product (GDP) refers to the market value of all officially recognized final goods and services produced within a country in a given period. GDP per capita is often considered an indicator of a country's standard of living.*" Moreover, tendencies in GDP time-series data are used for detecting recession periods and periods of growth. GDP data series are the most frequently requested data in many databases (for example see [8]). GDP data are usually available as annual and quarterly time series; in order to have sufficient number of observations, we used the quarterly GDP time series.

The selection of countries for this analysis was based on the definition of GDP: GDP is equal to the sum of all private consumption, all government spending, all country business spending on capital, and the nation's total net exports, calculated as total exports minus total imports [1]. As the Czech economy is bound with other economies by its trade, we selected four major export destinations, as listed in the "Main External Trade Partners, January - February 2012" [6]: Germany, Poland, Slovakia and France. We expected that the development of the main economic indicators – GDPs in this case – in these countries would have an influence on the development of this indicator in the Czech Republic. Moreover, we wanted to decide the causality in this development.

As a tool to compare the development trends, we considered multivariate time-series models applied to time series of the gross domestic product of the selected countries – namely the vector autoregression model (VAR) and the vector error correction model (VECM). Moreover, the analysis tried to reveal an impact of the main trading partners on the Czech economy. This meant to check if the widely used argument for a positive impact of the German economy on the Czech economy (based mainly on the increased production of car components and their export to Germany) is true. We also checked if the expectation of a negative influence of the Polish economy on the Czech economy is true as well. The latter expectation is based on the fact that cheaper Polish agricultural imports cause a decrease in the Czech agricultural production. Both expectations are tested using bivariate analysis.

The article is organized as follows: The next section is devoted to the data description and basic unit root tests of the selected data. The third part describes the models used and the results obtained. A conclusion ends the paper.

[1] Silesian University in Opava/School of Business Administration in Karvina, Department of Mathematical Methods in Economics , Univerzitní náměstí 1934/3, 733 40 Karviná, the Czech Republic, tosenovskyfilip@opf.slu.cz.
[2] 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, the Czech Republic, mielcova@opf.slu.cz.

2 Description of selected GDP data

For our model, we considered five GDP time series – quarterly data from 1:1995 to 4: 2011 from the Czech Republic, Germany, Poland, Slovakia and France. The data are available as the official OECD statistics [8]. All the data are seasonally adjusted. For illustration, the seasonally adjusted data are given in Figure 1.

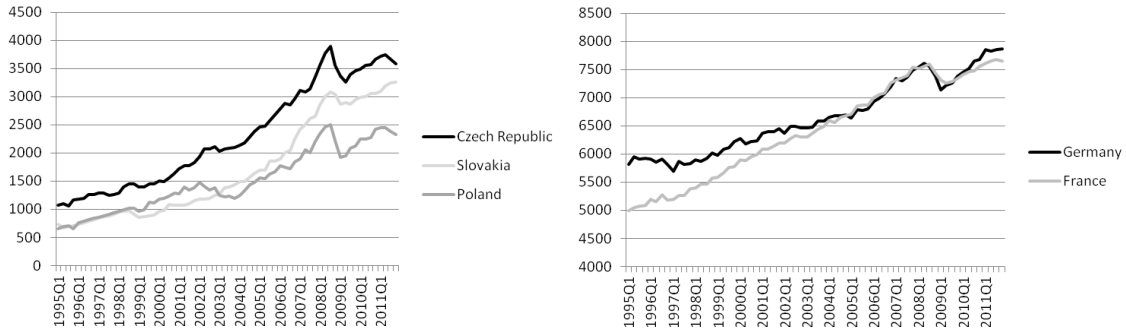


Figure 1 GDP data seasonally adjusted. Source: Calculations, gretl program.

We intended to use either a vector autoregression model or a vector error correction model to investigate the relationship among the GDP time series of the five countries with intertwined international trade. Thus, we needed first to test for stationarity of the time series. Here we used the Dickey-Fuller test for each individual time series, assuming that the true process follows in each case a simple one-lag regression with a generally nonzero constant term [2, 5]. Thus, the test was based on the estimation of the equation:

$$y_t = \beta_0 + \rho y_{t-1} + \varepsilon_t \tag{1}$$

The test checks whether the coefficient ρ is different from 1. Equation (1) can be rearranged such that:

$$y_t - y_{t-1} = \Delta y_t = \beta_0 + (\rho - 1) y_{t-1} + \varepsilon_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t \tag{2}$$

If the coefficient β_1 is not statistically different from 0, then ρ is not statistically different from 1. Results of the estimations of (2) on all five data sets are given in Table 1. None of expected β_1 is statistically significant, and so the presence of the unit root in all the data series cannot be rejected.

	β_0	β_1
Czech Republic	44.1542	-0.0030 (0.1317)
Germany	-9.0090	-0.0060 (0.7029)
Poland	44.5706	-0.0131 (0.4965)
Slovakia	16.1820	0.0129 (0.1329)
France	92.4534	-0.0082 (0.3437)

Table 1 Results of Dickey-Fuller tests estimation based on equation (2), p-values for the coefficient β_1 are given in parentheses. Source: Own calculations.

3 Multivariate models of GDP data from selected countries

The main aim of this analysis is to determine how changes in GDP of one country affect changes in GDP of another country. The first step is to decide which of the models to use. The VAR model is appropriate for not cointegrated time series, while the VECM can be used under the condition of cointegration in the data series.

In order to determine whether there is a cointegration relationship between the variables, the Johansen test was performed. The test is based on the VECM representation of the form:

$$\Delta X_t = \Pi X_{t-k} + \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + \ldots + \Gamma_{k-1} \Delta X_{t-(k-1)} + \varepsilon_t \tag{3}$$

where X_t is a 5×1 vector containing the GDP variables for all five countries, and Π and Γ are 5×5 matrices of coefficients.

The test for cointegration is calculated by studying the rank of the Π matrix via its eigenvalues. The length of lag in the model can be determined using different information criteria; the Schwartz Bayesian criterion and

the Hannan-Quinn criterion indicate the lag length of 1. However, the Akaike information criterion indicates the length of 10 lags. The Johansen trace test generally reveals that no cointegration relation between the variables can be expected; values of Johansen cointegration trace tests for three different lag numbers are given in Table 2.

r (number of cointegrating vectors under the H_0)	Test statistics (1 lag)	Test statistics (4 lags)	Test statistics (10 lags)
0	289.03	94.475	320.25
	(0.0000)	(0.0001)	(0.0000)
1	196.12	53.701	154.16
	(0.0000)	(0.0115)	(0.0000)
2	121.20	32.014	48.108
	(0.0000)	(0.0267)	(0.0001)
3	63.465	16.293	15.022
	(0.0000)	(0.0362)	(0.0574)
4	20.398	6.5465	0.61347
	(0.0000)	(0.0105)	(0.4335)

Table 2 Johanson test for cointegration (trace test), p-values are in given in parentheses. Source: Calculations, gretl program.

Thus, the standard VAR model in differences should be performed. The selected VAR model is based on the estimation of the equation system of the form:

$$\Delta X_t = c + \phi_1 \Delta X_{t-1} + \phi_2 \Delta X_{t-2} + \ldots + \phi_p \Delta X_{t-p} + \varepsilon_t \tag{4}$$

where X_t is a 5×1 vector containing the GDP variables for all five countries, c is a 5×1 vector of coefficients and ϕ are 5×5 matrices of coefficients. The length of lag in the model, determined by the Schwartz Bayesian criterion and the Hannan-Quinn criterion was set to be of the length of 1. Results of the estimation are in Table 2.

GDP	Δ CR	Δ Germany	Δ Slovakia	Δ Poland	Δ France
constant	21.7192	24.2548	20.9615	10.6158	42.8637
	(0.0969)*	(0.0652)*	(0.0216)**	(0.4100)	(0,0000)***
Δ CR(t-1)	0.0682	0.2147	0.1017	0.1943	0.1186
	(0.7072)	(0.2403)	(0.4178)	(0.2832)	(0.3840)
Δ Germany(t-1)	0.2418	-0.1345	0.0559	0.4556	0.2214
	(0.1391)	(0.4088)	(0.6173)	(0.0061)***	(0.0719)*
Δ Slovakia(t-1)	0.1512	-0.1897	0.1335	0.0227	-0.0359
	(0.4353)	(0.3295)	(0.3190)	(0.9059)	(0.8041)
Δ Poland(t-1)	0.3193	0.1366	0.0664	-0.0446	0.0851
	(0.0584)*	(0.4134)	(0.5628)	(0.7869)	(0.4947)
Δ France(t-1)	-0.2042	0.1014	0.1400	-0.1749	-0.3861
	(0.3686)	(0.6554)	(0.3716)	(0.437)	(0.0258)**

Table 3 Results VAR estimation based on equation (4), p-values for the coefficients are given in parentheses. (Significance levels: * 0.1 level, **0.05 level, *** 0.01 level) Source: Own calculations.

The results indicate that there is, for instance, a positive influence of the Polish economy on the Czech economy as well as a positive influence of the German economy on the economy of France and Poland. However, most of estimated coefficients are not statistically significant.

The Czech versus the German Economy

The above model indicates that there is no significant influence between the Czech and the German economies, taking into account multivariate analysis based on several trading partners' economies. However, we can test the possible long-term and short-term influence using comparison of the two data series, those of the Czech Republic and Germany. In order to test the possible cointegration between the two time series and to detect the possible long-term relation we have chosen the Granger-Engle cointegration test [3]. Thus the result of the ordinary least squares estimation:

$$GDP_{CR} = \text{-}6871.90 + 1.3760 GDP_{Germany} \qquad R^2 = 0.9688 \qquad \text{DW} = 0.4235 \tag{5}$$

The Dickey-Fuller test on the residuals of the model indicates that there is the long-term cointegrating relationship:

$$\Delta res^{(4)} = -3.8230 - 0.1859 res^{(4)}_{(t-1)} \qquad \text{coefficient st. error} = 0.0810 \qquad \text{p - value} = 0.0250 \ ** \tag{6}$$

The result indicates that the two variables have a long-term relationship. As we can expect the cointegration between the two variables, the VECM is appropriate to describe the situation. The number of lags in the model was set to be 2 as was indicated by the Hannan-Quinn criterion. The respective VECM results are:

$$\begin{aligned} \Delta GDP_{CR} = 16.4315 + 0.5092.\Delta GDP_{Germany} + 0.2330.\Delta GDP_{Germany,(t-1)} - 0.1122 res^{(4)}_{(t-1)} \qquad R^2 = 0.3366 \\ (0.0000)*** \qquad (0.0595)* \qquad (0.0806)* \qquad \text{DW} = 1.5397 \end{aligned} \tag{7}$$

The error-correction term coefficient is statistically significant and negative. That means that if a unite shock appears, the 11.22% of the shock will be absorbed in the next period, and it will be completely absorbed after nine periods.

To study the short term causality, we have chosen the Granger causality test (2 lags). Results indicate that:

- The change in the Czech GDP Granger causes the change in the German GDP (F=7.2927 (p=0.0015))
- The change in the German GDP does not Granger cause the change in the German GDP (F= 1.5881 (p= 0.2128))

This result contradicts to the expected influence of the German economy to the Czech economy. The respective impulse – response functions are given in Figure 2.

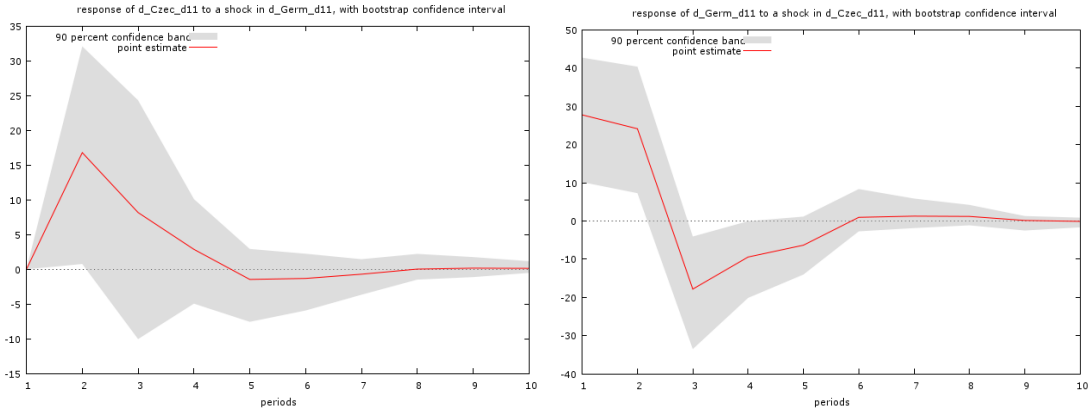


Figure 2 Impulse-response functions of the Czech and German GDP change on the unite shock in the German and Czech GDP change, respectively. Source: Calculations, gretl program.

The Czech versus the Polish Economy

Similar analysis as above is to be performed on the Czech and the Polish time-series data. In order to test the possible cointegration and long-term relationship we have perform the Granger-Engle cointegration test. The result of the ordinary least squares estimation do not confirm the expectation of negative impact of the Polish economy:

$$GDP_{CR} = -143.759 + 1.6260 GDP_{Poland} \qquad R^2 = 0.9661 \qquad \text{DW} = 0.0.3360 \tag{8}$$

The Dickey-Fuller test on the residuals of the model indicates that there is the long-term cointegrating relationship:

$$\Delta res^{(7)} = -3.2766 - 0.1748 res^{(7)}_{(t-1)} \qquad \text{coefficient st. error} = 0.0686 \qquad \text{p - value} = 0.0132 \ ** \tag{9}$$

The result indicates that the two variables have a long-term relationship. The VECM is appropriate to describe the situation [7]; number of lags in the model was set to be 1 as was indicated by the Hannan-Quinn criterion as well as the Schwartz Bayesian criterion. The respective VECM results are:

$$\begin{aligned} \Delta GDP_{CR} = 19.1368 + 0.7144.\Delta GDP_{Poland} - 0.2254 res^{(4)}_{(t-1)} \qquad R^2 = 0.5657 \\ (0.0000)*** \qquad (0.0030)*** \qquad \text{DW} = 1.7541 \end{aligned} \tag{10}$$

The error-correction term coefficient is statistically significant and negative. That means that if a unite shock appears, the 22.54% of the shock will be absorbed in the next period, and it will be completely absorbed after four periods.

To study the short-term causality we performed the Granger causality test. Results indicate that (2 lags):

- The change in the Czech GDP Granger causes the change in the GDP of Poland (F= 6.8103 (p=0.0022))
- The change in the Polish GDP does not Granger cause the change in the German GDP (F= 3.0221 (p= 0.0562))

The results contradict to the expected implication about the influence of the economy of Poland on the economy of the Czech Republic. The impulse-responses functions are given in Figure 3.

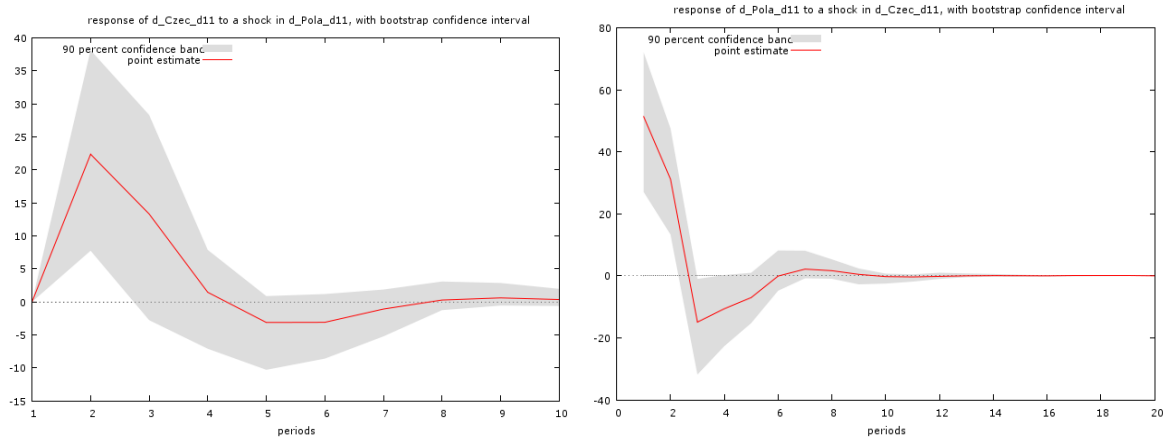


Figure 3 Impulse-response functions of the Czech and Polish GDP change on the unite shock in the Polish and Czech GDP change, respectively. Source: Calculations, gretl program.

4 Conclusions

This article compares development of the Czech economics with development of economics of several countries in the European Union for the period of 1995-2011 measured in terms of the gross domestic products of selected countries. The data series are I(1) stationary, and has no expected cointegration as a whole. The respective VAR model indicates that there is a positive influence of the Polish economy on the Czech economy as well as the positive influence of the German economy on the economy of France and Poland. However, there are no signs of expected influence of the German economy on the Czech economy.

Taking into account the pure pair wise comparisons of the Czech economy with those of Poland and Germany, the results reveal long-term cointegrating relationship, described by the estimated VECM models. The short-run relationships were tested by the Granger causality test. The results do not support the theory of the influence of the German and Polish economy on the Czech economy. While the change in the Polish and German GDP does not cause the change in the German GDP, the change in the Czech GDP precedes the change in the German and Polish GDP.

References

[1] Abel, A. B., Bernanke, B. S.: *Macroeconomics (5th ed.).* Pearson Addison Wesley, 2005.
[2] Arlt, J.: *Moderní metody modelování ekonomických časových řad.* Grada, Prague, 1999.
[3] Engle, R. F., Granger C. W.: Cointegration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, **55** (March 1987), 251-276.
[4] Gross domestic product. In: *Wikipedia*: *the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- [accessed: 2012-05-18]. From: http://en.wikipedia.org/wiki/Gross_domestic_product.
[5] Hamilton J. D.: *Time Series Analysis.* Princeton University Press, Princeton, New Jersey, 1994.
[6] Main External Trade Partners, January - March 2012. CSU. *Czech Statistical Office* [online]. [accessed: 2012-05-18]. From: http://www.czso.cz/csu/2012edicniplan.nsf/t/F400389B4F/$File/6001120304a.pdf
[7] Nielsen, H. B. *Non-Stationary Time Series, Cointegration and Spurious Regression* [online]. [accessed: 2012-05-18]. From: http://www.econ.ku.dk/metrics/Econometrics2_05_II/Slides/10_cointegration_2pp.pdf.
[8] OECD. *Organisation for Economic Co-operation and Development*: *OECD.Stat Extracts* [online]. [accessed: 2012-05-18]. From: http://stats.oecd.org/.

Managing monetary policy with fuzzy control

Tran Van Quang[1], Jarmila Radová[2]

Abstract. Most central banks in developed economies conduct monetary policy with inflation targeting regime. In this framework, their primary objective is to maintain price stability and other objectives of economic policies are supported only when the primary one is not compromised. To achieve their goals, the central banks have to manipulate the key monetary policy interest rate. Usually, the decision on setting this interest rate is claimed to be based on the output of some sophisticated model. Nevertheless, the recent experience has shown that it is not always the case and interest rate changes are still heavily relied on the judgements of central bankers. Then one should ask how the interest rate is set. And one possible answer to this question is the use of fuzzy control to appropriately quantify it. This approach then is verified with the decisions on 2W reporate of the Czech Central Bank from period 2000 up to now.

Keywords: Monetary policy, 2W repo rate, Fuzzy control

JEL classification: E52, E58

1 Introduction

The main objective of most independent central banks today is to keep the price level in the economy stable. Besides this principle goal, they also have other objectives which are: to support economic growth and to keep high employment, but these goals are pursued just in case they are reconcilable with the main objective. The monetary policy is then executed through setting the key monetary policy interest rate. Though every self-respected central bank maintains that its decisions on interest rate setting are (DSGE) model based, in reality its decisions still are very esoteric. As a result, nobody knows what is the rule on which the final decision on interest rate relies. To overcome this problem, we suggest the fuzzy control approach whose essence is as follows. From the regulation point of view, monetary policy in the inflation targeting framework is a traditional control problem. In this framework, first, the inflation target is publicly set and then the central bank will manipulate the interest rate in such a way in order to steer the real inflation in the economy to its targeted value. Since all the underlying processes are not fully known, fuzzy control can be an alternative applicable to conducting monetary policy with inflation targeting regime. Fuzzy control converts correct but vague expert knowledge to crisp values and has succeeded in various technical areas. Hence it is worth being applied to control monetary policy and we will explain how this control approach can be used for this purpose. Our paper is structured as follows: the next section is devoted to the basics of fuzzy control. After that, we will show how fuzzy control is applied to monetary policy. Then we will verify to show how monetary policy would have looked like with the fuzzy control technique and it will be compared to how it has been actually run by the Czech National Bank for the last decade in the Czech Republic. Finally, concluding remarks will be made in the last section.

2 Fuzzy logic and fuzzy control

As it has been mentioned, monetary policy in the inflation targeting regime is an optimal control problem. In many real-life cases, if the problem is mathematically well defined and precisely formulated, then the optimal control strategy is the solution of optimization problem and the optimal controls can be found. But it is not always the case. In such situations, one can only rely on the experience of experts which is generally correct, but not easily quantified in some exact formula. For example, if we want to control the temperature inside of a room, but for some reason, we do not know the corresponding thermodynamic process, then the temperature can be regulated as follows. If it is hot in the room, then we should open the window or turn on the air-conditioner in the room to lower the temperature. On the contrary, if it is cold, then we should close the window or turn off the air-conditioner.

[1]University of Economics in Prague, Department of Statistics and Probability, nám. W. Churchilla 4, Praha 3, e-mail: tran@vse.cz
[2]University of Economics in Prague, Department of Banking and Insurance, nám. W. Churchilla 4, Praha 3, e-mail: radova@vse.cz

Further, if it is too hot inside, then the window must be widely opened, if it is not so hot, then just ajar and so on. The control is expressed in a form of natural language, which, on the one hand, is very easy to understand, but on the other hand, it is quite vague. Fuzzy control is a control method which changes these vague rules into an exact control strategy. A fuzzy control process dealing with this issue is divided into four parts. First, the real inputs should be changed into fuzzy values. Then, we choose a set of rules expressing common knowledge on what should be done in order to achieve the goal. After that, we should combine these sometimes contradicting rules in order to choose the best action to be done to reach the goal of the controlled process. Finally, we have to convert the fuzzy values back to real values of the controlled process, which is called defuzzification to set an exact control procedure.

Fuzzification

Fuzzification is defined as the mapping a value v of quantity V into interval [0.1], or formally $\mu : V \mapsto [0, 1]$. The mapping is called the membership function which measures the degree of membership of v to a certain set (category). Unlike a conventional set, where the degree of membership of any object in the set is either 0 or 1, the membership function can be Z-shaped, triangular or trapezoidal and it can attain any value between 0 and 1. For fuzzy sets (A and B), the following operators are defined:

- the fuzzy intersection operator $\wedge$ (AND connective):

$$\mu(A \wedge B) = \min\{\mu(A), \mu(B)\}, \tag{1}$$

- the fuzzy union operator $\vee$ (OR connective):

$$\mu(A \vee B) = \max\{\mu(A), \mu(B)\}, \tag{2}$$

- the fuzzy complement (not operation):

$$\mu(\text{not}A) = 1 - \mu(A). \tag{3}$$

Fuzzy rules

Fuzzy rules come from the experience of experts and include a series of IF - THEN statements which are characteristic to human thinking and perception. Each if-then statement with an antecedent and a consequent forms a fuzzy proposition. The antecedent can also contain a combination of propositions connected by logical operators AND and OR. For example, a rule can look like this:

IF v_1 is A_1 and ... and v_i is A_i, THEN Do B_1.

Rules combination

Since there is a set of experts' rules, we have to determine the grades of fulfilment of each rule. Usually, we evaluate the consequent of each individual rule by using the minimum implication, but sometime the product implication is also used. Finally, all partial individual consequences are aggregated into an overall result.

Defuzzification

As the outputs from the previous part are fuzzy numbers presenting the final consequent from a set of rules of form if - then. These fuzzy numbers need to be converted back to a crisp number as the value of the control which should be set. There are many ways to do so, but the most often one is the so called center of gravity method (COG). The sought crisp value is the centroid of the area under the graph of the membership function of the output set (see Figure **??**). Formally, the center of gravity is defined as average value as follows:

$$i_B = \frac{\int_U u\mu_\beta(u)du}{\int_U \mu_\beta(u)du}. \tag{4}$$

3 Application of fuzzy control to monetary policy conduction

In this section, the fuzzy control principle presented in the previous part is applied to managing monetary policy of the Czech National Bank. In the Czech Republic, according to its constitution and Act No. 6/1993 Coll. on the Czech National Bank, monetary policy is mandated to the care of the Czech National Bank (CNB). The primary CNB's objective is to maintain price stability in the Czech economy. At the same time, the CNB may support the general economic policies of the Czech government, basically in terms of economic growth, if its primary goal is not compromised. To achieve its objectives, the CNB have been using inflation targeting regime since 1998. In this regime, the CNB makes an explicit public announcement on the inflation target it wants to reach. To achieve this goal, the CNB has to manipulate its key monetary policy interest rate 2W repo rate[1] according to the development of the Czech economy in order to keep the real inflation rate as close to its preset target as possible. The CNB can also support the growth of the Czech economy by using this tool if the primary goal is not affected. So far the decisions on the desired level of this rate have been made by the CNB's Board and they claim that these decisions are based on the output of macroeconomic modelling as well as on their judgements on the future development of the economy which is totally undisputable, but at the same time still very vague.

The CNB's monetary policy as described above is clearly a typical control problem. The CNB has an inflation target, it uses its 2W repo interest rate as a tool to drive the real inflation in the economy to its target value. Since it does not exactly know the nature of all underlying economic processes, the CNB cannot rely only on the model output and it has to take into account the experts' knowledge which is not unquantifiable. But the fuzzy control approach can helps. In this case, the CNB's monetary policy has two objectives: to keep the inflation level stable, for example, around 2% and to support economic growth at a sustainable rate while the inflation objective is superior to supporting economic growth. We assume that the sustainable growth rate is the long-term growth rate which is the average of a long period. Using fuzzy control approach, first the crisp economic data on inflation and economic growth available to the CNB has to be converted to fuzzy data via their corresponding membership functions.

If the inflation target is 2%, then we can define inflation is just right meaning the membership value is 1 when it is in interval 1,5 - 2,5%. The grade of just rightness decreases toward value 0 when it falls or grows towards 0.5 or 3.5 respectively. Inflation is not just right definitely when it is out of interval (0.5,3.5). Similarly, inflation is definitely low when it is below 0.5. The grade of lowness of inflation is decreasing toward 0 when inflation rises towards 1.5% and it is not low any more if it is higher than 1.5% (membership value is 0). As far as the highness of inflation is concerned, it is not high if it is less than 2.5% (membership value is 0). Then the grade of highness increases when it rises towards 3.5%. And the inflation is surely high (membership value is 1) if it is higher than 3.5%. The membership function of inflation is shown in Figure 1 (the left one).

By the same token, we construct the membership function of economic growth. In this case the average growth

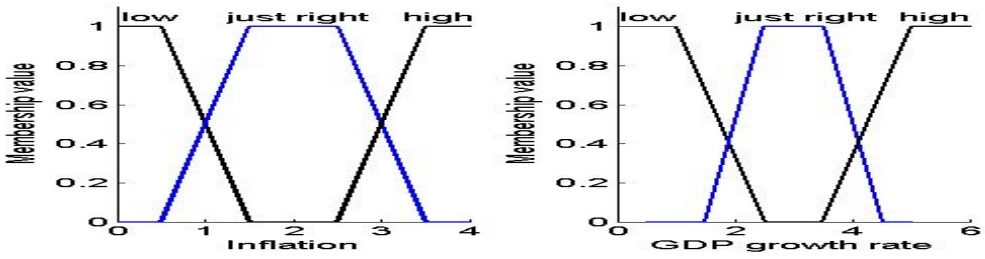


Figure 1 The membership function for inflation and GDP growth rate

rate from the ten-year period 2001 - 2011 which is 3.0% is considered to be the sustainable rate The membership function of economic growth rate is shown in Figure 1 (the right one). With these two membership functions, the crisp inputs in terms of inflation and economic growth rate can be fuzzified. Now, as the fuzzy control method suggests, the fuzzy rules are needed. In accordance with monetary theory [4] and with respect to the objectives of the CNB, the following rules can be constructed:

- IF inflation is low and GDP growth rate is also low, THEN the CNB reduces the 2W repo rate to support economic growth because the lower repo rate leads to lower other market interest rates via transmission mechanism which discourages savings and encourages spending and eventually leads to higher output without compromising inflation target.

[1] 2W repo rate is the rate the CNB uses to withdraw the excess liquidity from the market

- IF inflation is low and GDP growth rate is just right, THEN the CNB reduces the 2W repo rate to support economic growth
- IF inflation is just right and GDP growth rate is low , THEN the CNB reduces the 2W repo rate to support economic growth
- IF inflation is high and GDP growth rate is also high, THEN the CNB increases the 2W repo rate to lower inflation because the higher repo rate leads to higher other market interest rates via transmission mechanism which discourages spending and encourages savings and eventually leads to lower inflation to meet the inflation target.
- IF inflation is high and GDP growth rate is just right, THEN the CNB reduces the 2W repo rate to meet its inflation target because price stability objective is superior to economic growth objective
- IF inflation is just right and GDP growth rate is high, THEN the CNB keeps the 2W repo rate unchanged
- IF inflation is just right and GDP growth rate is also just right, THEN the the CNB keeps the 2W repo rate unchanged
- IF inflation is low and GDP growth rate is high, THEN the CNB keeps the 2W repo rate unchanged
- IF inflation is high and GDP growth rate is low, THEN the CNB keeps the 2W repo rate unchanged because low economic growth will amend the inflation rate to the lower level.

These rules are summarized in Table 1. From Table 1 it is clear that the CNB can lower its key interest rate either

Table 1 The set of monetary policy rules

	Low Inflation (IL)	Just right Inflation (IJR)	High Inflation (IH)
Low Production (GL)	decrease	decrease	do not change
Just right Production (GJR)	decrease	do not change	increase
High Production (GH)	do not change	do not change	increase

when inflation and growth rate are low or when inflation is low and growth rate is just right or when inflation is just right and growth rate is low. On the other hand, the CNB should increase its key interest rate either when inflation is high and growth rate is also high or when inflation is high and growth rate is just right. Finally it does not change its monetary interest rate when it does not have to increase and to decrease it.
For completeness, we also build a set of rule for the case when the CNB is an inflation fighter. In this case, the CNB will care only about inflation and increases its repo rate if inflation is high, decreases the repo rate if inflation is low and keep it unchanged when the inflation is just right regardless of the development of the economic development.

Fuzzy interference

Since there are nine monetary policy rules and inflation and GDP growth rate can fall into one of these possibilities. The final rule is the combination of possibilities required the same action from the CNB connected by operator OR. For operator OR, the minimum rule is used to get an output fuzzy number. As we see, the CNB decreases the interest rate when either inflation and economic growth rate are low or inflation is just right and growth rate is low or inflation is low and growth rate is just right, hence:

$$\text{DECR} = (\text{IL} \wedge \text{GL}) \vee (\text{IL} \wedge \text{GJR}) \vee (\text{IJR} \wedge \text{GL}).$$

Similarly, the increase of the interest rate happens when either inflation and growth rate are high or when inflation is high and growth rate is just right.

$$\text{INCR} = (\text{IH} \wedge \text{GH}) \vee (\text{IH} \wedge \text{GJR}).$$

And the interest rate is kept unchanged when both previous cases do not happen, which means:

$$\text{UNCH} = 1 - \text{DECR} - \text{INCR}.$$

Defuzzification

The output fuzzy number from the previous part needs to be converted to a crisp number. In accordance with the changes of the repo rate taken by the CNB so far, we define the membership function for the interest rate change as follows. Since the size of the biggest change so far is 0.75%, the membership for sure decrease or increase is set for interval (0.5, 1%) respectively. Then it decreases towards 0. Keeping the interest rate unchanged ranges from -0.25 to 0.25% with the just right value equal 0 (see Figure 2, upper part). The resulting crisp value then is calculated by the center of gravity method (COG) proposed by Mamdani [1] which is the value of the ratio of the abscissa of the centre of gravity of the area in blue to the area covered by the membership function. For example, for the inflation = 1,1% and the growth rate = 1,8%, we calculate the membership values for inflation and growth as follows: μ_I^l = 8/30, μ_I^{jr} = 0.1 and μ_I^h = 0 and μ_G^l = 4/30, μ_G^{jr} = 0.2 and μ_G^h = 0. Using the experts rules and fuzzy interference we get the following membership values of the interest change: μ_{DECR} = 0.2, μ_{INCR} = 0 and μ_{UNCH} = 0.8. This values are then defuzzified and the crisp value of interest rate change is roughly 0,2 (see Figure 2, upper part). Since the CNB changes its 2W repo rate in multiples of a quarter percent, in this case the CNB lowers its key interest rate by one quarter percent if the rate is not zero already.

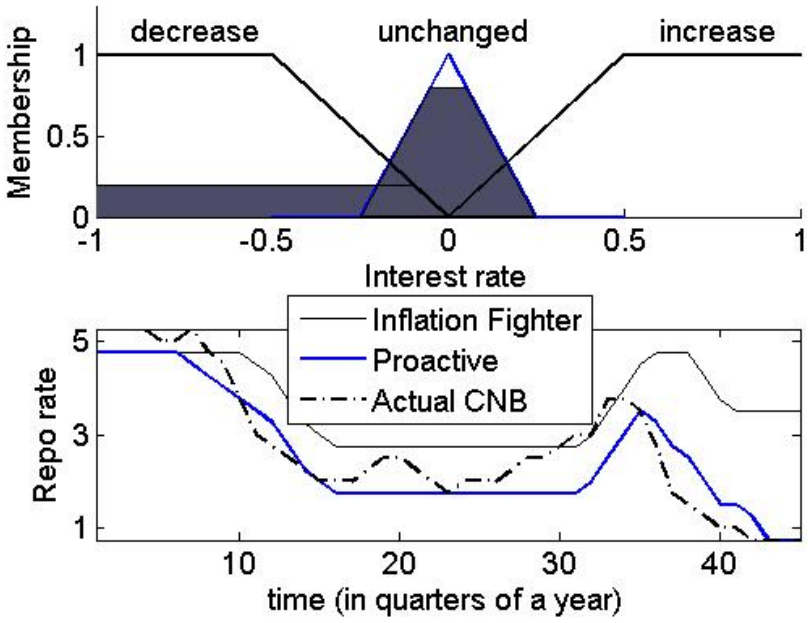


Figure 2 Membership function for interest rate change and verification results

4 Verification of the CNB's monetary policy by fuzzy control

In this part, the fuzzy control approach is used to verify the monetary policy having been conducted by the CNB from 2000 to the end of 2011. For this purpose, we use the data on quarterly inflation and GDP growth rate published by the Czech Statistical Office [6]. To determine the membership values of inflation and GDP growth rate, the data from each quarter is used for the same quarter. The reason for it is that though this data was not available to the CNB yet, but at those moments it could be well aware of their trends and could forecast them with reasonable accuracy. As far as the inflation target of the CNB is concerned, since their values were changed over the course of time, we have to adjust them correspondingly. The CNB's inflation targets are set as follows [5]: first, until the end of 2000 the target was set as a band from 4 to 5% and the price stability was measured in terms of pure inflation. In 2001 the target band was extended to 3,5 to 5,5%. Since the beginning of 2002, the price stability had been measured by the Consumer Price Index (CPI) and the target interval was 3 - 5%. This target interval was gradually reduced to 2 - 4% towards the end of year 2005. From the beginning of 2006 the target interval was changed to a point target value a the value was 3% for period 2006 - 2009. Since the beginning of 2010 this value is 2%. As we have only data on Czech CPI for the whole period, 1% is added to the target value when it is set in pure inflation term. When determining the membership values, 0,5% is always added to both sides of the

target values or target bands for the just right case. Otherwise the membership functions are constructed as shown in Figure 1. Otherwise, the remainder of the procedure is the same as it was in the previous section. Finally, when calculating the size of changes of the key CNB's interest rate, we proceed as follows: we calculate only once for a quarter while the CNB could change it several times in a quarter (and actually it did happen four times during the examined period). The size of a change is 0 if it is less than 0.125%, it will be 0.25% if it is in interval (0.125, 0.375) a it is 0.5% if it is greater than 0.375% since the CNB changes its interest rate in multiples of a quarter of a percent if it does change. The calculation was carried out in the Matlab environment and is available on request.
The result of the verification is shown in Figure 2 (lower part). We reexamine two possible policies. One is called as the proactive policy which is in line with the official CNB's policy (the lower and thicker line in the lower part of Figure 2), the second one is dubbed as the inflation fighter policy (the upper and thinner line in the lower part of Figure 2). The actual interest rate policy is the dotted line in the lower part of Figure 2. Visually inspecting this figure, it is clear that both examined policies can capture the development of the Czech economy and have similar course as the actual CNB's policy. If the emphasis is put on inflation fighting, the the interest rate are higher while the proactive policy does not substantially differ from the actual one. The differences might result from the forecast available to the CNB when they had to make decisions on interest rate or they might be of operational pre-emptive origin.

5 Conclusion

In this paper we have applied fuzzy control approach to monetary policy in the inflation targeting regime. As the underlying mechanism controlling economic activities are not exactly quantifiable, running monetary policy by a central bank still is an intuitive matter rather a mathematically well defined control problem. As such, fuzzy control seems to be a suitable approach for this purpose. We have used the simplest model of a central bank in inflation targeting regime, a fuzzy control system is capable of converting well-accepted qualitative monetary policy principles to quantitative response in terms of the size of the monetary policy interest rate of a central bank. The fuzzy control approach is on one hand very strict, but on the other hand, is very flexible in terms of those coefficients in the widely used Taylor [3] rule and makes it a special case of fuzzy control approach due to its nonlinear a nonparametric nature. Since we have used a model-free approach to quantify the decisions of the CNB on interest rate, it would be interesting to combine this technique with a DSGE model to examine whether it can improve the outputs of the often used DSGE model.

Acknowledgements

The authors are grateful for the financial support of Czech Science Foundation No. P402/12/G097.

References

[1] Mamdani, E.H., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies* **7** (1975), 1 - 13.

[2] Novak, V.: *Základy fuzzy modelování*. BEN Technická literatura, Praha, 2000.

[3] Taylor, J. B.: Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy* **3** (1993), North-Holland, 195-214.

[4] Walsh, C.E.: *Monetary Theory and Policy*, MIT Press, Massachusetts, 2010.

[5] www.cnb.cz

[6] www.czso.cz

Two-dimensional voting bodies: The case of European Parliament

František Turnovec[1]

Abstract. By a two-dimensional voting body we mean the following: the body is elected in several regional voting districts by proportional system based on multi-party competition of national political parties. Then the members of the body exercise dual responsibility: responsibility following from the party membership and responsibility following from regional affiliation.

In this paper we formulate the following problem: Taking as decisional units national chapters of European political parties, is there a difference between a priori voting power of national groups in the case of "national" coordination of voting and in the case of "partisan" coordination of voting? By coordination of voting we mean two step process: in the first step there is an internal voting in the groups of units (national or partisan), in the second step there is a voting coordination of aggregated groups (European political parties or national representations). In the both cases the voting has an ideological dimension (elementary unit is a national party group), difference is only in dimension of aggregation. Power indices methodology is used to evaluate voting power of national party groups, European political parties and national representations in the cases of partisan and national coordination of voting behaviour.

Keywords: European Parliament, European political parties, ideological coordination, national coordination, Shapley-Shubik power index, two-dimensional voting bodies, voting power of national party groups

JEL Classification: D71, C71
AMS Classification: 91F10

1 Introduction

By a two-dimensional voting body we mean the following: the body is elected in several regional voting districts by proportional system based on multi-party competition of national political parties. Then the members of the body exercise dual responsibility: responsibility following from the party membership and responsibility following from regional affiliation.

There exist more examples of two-dimensional voting bodies (committees). Practically all national parliaments have in some sense two-dimensionality features, especially upper houses in bi-cameral systems. Their individual members represent citizens of the region they were elected in and on the other hand they are affiliated to some political party. One of the voting bodies clearly exhibiting two-dimensional face is the European Parliament.

Increasing number of studies are focusing attention to constitutional analysis of European Union institutions and distribution of intra-institutional and inter-institutional influence in the European Union decision making. Most of the studies are related to distribution of voting power in the EU Council of Ministers as reflecting the influence of member states (or, more precisely, member states governments). Significantly less attention is paid to the analysis of European Parliament. It is usually studied in one-dimensional (partisan) framework, taking as a basic decision making unit European political party.

In this paper we extend Nurmi [5] and Mercik, Turnovec, and Mazurkiewicz [4] analysis and formulate the following problem: Taking as decisional units national chapters of European political parties, is there a difference between a priori voting power of national groups in the case of "national" coordination of voting and in the case of "partisan" coordination of voting? By coordination of voting we mean two step process: in the first step there is an internal voting in the groups of units (national or partisan), in the second step there is a voting coordination of aggregated groups (European political parties or national representations). In the both cases the voting has an ideological dimension (elementary unit is a national party group), difference is only in dimension

[1] Charles University in Prague, Faculty of Social Sciences, Institute of Economic Studies, Opletalova 26, 110 00 Praha 1, Czech Republic, turnovec@fsv.cuni.cz.

of aggregation. Power indices methodology (Shapley and Shubik [6]) is used to evaluate voting power of national party groups, European political parties and national representations in the cases of partisan and national coordination of voting behaviour.

EP is institutionally structured on ideological principle, the individual members (national party groups) work in factions of the European political parties. Hix, Nouri and Roland [1] demonstrated, using empirical data about voting acts in EP of the fifth term, that while ideological dimension in EP voting prevails (in almost 80% of cases EP members voted according European party affiliation), there were still more than 20% of voting driven by national dimension (voting by national affiliation). Consequently, to measure the influence in the EP, basic decision making unit is a national party groups and it makes sense to measure not only voting power of European political parties and/or voting power of national representations, but also the voting power of national party groups, both in ideologically driven voting and nationally driven voting. European political parties cohesion is lower than cohesion of their national chapters.

2 Two-level committee model of power decomposition

Let N = {1, 2, …, n} be a set of agents, [γ, $\boldsymbol{\omega}$], be a committee with quota γ and weights ω_i, i ∈ N, and $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_n)$ be the vector of Shaply-Shubik power indices of agents of the committee. Then π_i is a probability that agent i ∈ N will be in a pivotal situation.

Each agent i can be understood as a group G_i with cardinality ω_i (individual members of the committee belonging to i). Clearly $card(G_i) = \omega_i, \sum_{i \in N} card(G_i) = \tau$. Let $G_{ij} \in G_i$ be a subgroup j of the group G_i and ω_{ij} = card (G_{ij}), number of members belonging to G_{ij}. Assume that each group (agent) i is partitioned into m(i) subgroups G_{ij}. Then we can consider the following two step procedure of decision making: first each agent G_i looks for joint position in a subcommittee [γ_i; ω_{i1}, ω_{i2}, …, $\omega_{im(i)}$], where γ_i is the quota for voting in subcommittee i (e.g. the simple majority). There is a vote inside the group first (micro-game) and then the group is voting jointly in the committee on the basis of results of internal voting (macro-game).

If $\mathbf{p}(G_i) = (p_{i1}, p_{i2}, \ldots, p_{im(i)})$ is the power distribution in subcommittee G_i where p_{ij} be and internal power of subgroup G_{ij} in micro-game, then the voting power π_{ij} of the subgroup G_{ij} in macro-game is $\pi_{ij} = \pi_i p_{ij}$ expressing the probability of the subgroup G_{ij} being pivotal in the committee decision making. Using SS-power concepts it is easy to prove that

$$\sum_{j=1}^{m(i)} \pi_{ij} = \pi_i$$

so we obtained decomposition of the power of agent i among the subgroups G_{ij}.

3 Modeling distribution of power in European Parliament

To evaluate distribution of power of national party groups in European Parliament as basic decision making units we use the model of two-level committees from section 2. To reflect the double dimensionality in voting we use two dimensions of committee structure: the European party factions decomposed into national groups, and the national representations decomposed into the party groups. Basic unit remains the same in both cases: national party group. Then we obtain two schemes of decision making coordination: first based on European party factions and national party groups, second based on national representations and national party groups.

First (ideological) dimension leads to committee model A with European parties as agents voting together, [γ, p_1, p_2, …, p_n], the second (national) dimension leads to committee model B with national representations as agents voting together, [γ, n_1, n_2, …, n_m], where γ is the quota (the same for both models), p_i is the weight (number of seats) of European party i, n_k is the weight (number of seats) of member state k (n is the number of European parties, m is the number of member states).

Committee A generates n subcommittees A_j such that [γ_j, p_{1j}, p_{2j}, …, p_{mj}], where p_{kj} denotes number of members of party group j from country k, γ_j being a specific quota for subcommittee A_j. Each of these subcommittees consists of at most m national subgroups of the European political party j, where in each subcommittee the members of each party from the same member state k are voting together. We shall refer to the corresponding two-level model as the hierarchically structured committee {A/A_j}. Committee B generates m subcommittees B_k such that [δ_k, p_{k1}, p_{k2}, …, p_{kn}], where p_{ki} denotes number of members of party group j from country k, δ_k being a specific quota for subcommittee B_k. Each of these subcommittees consists of at most n party subgroups of the

national representation k, where in each subcommittee the members of from the same party j are voting together. We shall refer to the corresponding two-level model as the hierarchically structured committee $\{B/B_k\}$.

Let us denote by

α_j voting power of the European party *j* in the committee A (voting by ideological dimension), probability that party j will be pivotal in ideologically coordinated voting,

β_k voting power of the nation k in the committee B (voting by national dimension), probability that nation k will be pivotal in nationally coordinated voting,

α_{kj} voting power of the national segment k of party *j* in subcommittee A_j, probability that national segment k of party j will be pivotal in internal party voting,

βkj voting power of the national segment k of party *j* in subcommittee B_k, probability that party segment j of representation of country k will be pivotal in internal national voting,

π_{kj} voting power of the national segment k of party *j* in the committee $\{A/A_j\}$, probability that national segment k of party j will be pivotal in the grand committee voting based on ideological coordination,

φ_{kj} voting power of the national segment k of party *j* in the committee $\{B/B_k\}$, probability that party segment j of national representation k will be pivotal in the grand committee voting based on national coordination.

Using standard algorithms we can find SS-power indices α_j in committee A and α_{kj} in committees A_j (probabilities of being pivotal in corresponding committees) and then calculate $\pi_{kj} = \alpha_{kj}\alpha_j$ as conditional probability of two independent random events – pivotal position of j in grand committee A and pivotal position of k in subcommittee A_j. From probabilistic interpretation and properties of SS-power indices it follows that $\sum_{k=1}^{m} \pi_{kj} = \alpha_j \sum_{k=1}^{m} \alpha_{kj} = \alpha_j$. The sum of voting powers of national groups of European political party j in ideological voting is equal to the voting power of the European political party. The total power is decomposed among the national units of the party. In a more intuitive way: the national group k of political party j is in a pivotal position in compound committee $\{A/A_j\}$ if and only if it is in pivotal position in subcommittee A_j and the party j is in a pivotal position in committee A.

Less trivial is the following result: The country k is in a pivotal position in ideological coordination of voting if some party group from k is in pivotal position. Pivotal positions of national party groups of the same country in ideologically voting are mutually exclusive random events, hence the probability that some party group from state k is in a pivotal position is $\sum_{j=1}^{n} \pi_{kj} = \sum_{j=1}^{n} \alpha_j \alpha_{kj} = \theta_k$ (sum of power indices of all party groups from member state k). Then θ_k can be interpreted as a measure of country k influence in ideologically coordinated voting. From properties of SS-power it follows that $\sum_{k=1}^{m} \theta_k = \sum_{k=1}^{m} \sum_{j=1}^{n} \alpha_j \alpha_{kj} = \sum_{j=1}^{n} \alpha_j \sum_{k=1}^{m} \alpha_{kj} = \sum_{j=1}^{n} \alpha_j = 1.$ There is no other direct way how to evaluate θ_k.

In the same way we can find β_k in committee B and β_{kj} in committees B_k and then calculate $\varphi_{kj} = \beta_{kj}\beta_k$ as conditional probability of two independent random events - pivotal position of k in grand committee B and pivotal position of j in subcommittee B_k). Measure of party j influence in nationally coordinated voting is $\sum_{k=1}^{m} \varphi_{kj} = \sum_{k=1}^{m} \beta_k \beta_{kj} = \vartheta_j$ (sum of power indices of party group j from all member states).

4 Illustrative example

To illustrate methodology introduced above we use a simple hypothetical example.[2] Let us consider a parliament consisting of representatives of three regions A, B, and C decomposed into three super-regional parties L, M, R (altogether 9 regional party chapters of 3 super-regional parties). Distribution of seats is provided in Table 1. Entries in last row provide total number of seats of each party in the parliament, entries in the last column total number of seats of each region in the parliament, and entries in the main body of the table provide number of seats of each regional party chapter.

Regions	parties (seats)			
	L	M	R	total
A	7	10	3	20
B	15	15	0	30
C	3	22	25	50
Total	25	47	28	100

Table 1 Distribution of seats

Let us start with evaluation of distribution of power of super-regional parties in the parliament under assumption that regional chapters of each party are voting together. To do that we have to calculate power indices in the committee [51; 25, 47, 28].Total influence of super-regional parties in ideologically coordinated voting measured by Shapley-Shubik power indices (assuming simple majority quota): (1/3, 1/3, 1/3). Total influence of regional representations in regionally coordinated voting measured by Shapley-Shubik power indices (assuming simple majority quota) we have to calculate power indices in the committee [51; 20, 30, 50], voting power (1/6, 1/6, 2/3).

Influence of regional party chapters in ideologically coordinated voting:

Party group L: committee [13; 7, 15, 3]; voting power of regional party chapters of party L: (0, 1, 0). Total voting power of L in the parliament ideological voting equal to 1/3 is decomposed among the regional party chapters: (0, 1/3, 0).

Party group M: committee [24; 10, 15, 22]; voting power of regional party chapters of party M: (1/3, 1/3, 1/3). Total voting power of M in the parliament ideological voting 1/3 is decomposed among the regional party chapters: (1/9, 1/9, 1/9)

Party group R: committee [15; 3, 0, 25]; voting power of regional party chapters of party R: (0, 0, 1). Total voting power of R in the parliament ideological voting 1/3 is decomposed among the regional party chapters (0, 0, 1/3).

Evaluation of voting power of regional party chapters in ideologically coordinated voting is provided in Table 2. Entries in last row provide total voting power of each party in the parliament, entries in the last column total voting power of each region in the parliament in the case of ideologically coordinated voting, and entries in the main body of the table provide voting power of each regional party chapter.

Regions	parties (voting power)			
	L	M	R	Total
A	0	1/9	0	1/9
B	3/9	1/9	0	4/9
C	0	1/9	3/9	4/9
Total	3/9	3/9	3/9	1

Table 2 Decomposition of power in ideologically coordinated voting

[2] European Parliament elected in 2009 has 8 party groups, and decomposition might consist of 216 national party chapters; it is difficult to handle such a structure on limited space of conference proceedings. The results of empirical analysis for the European Parliament elected in 2009 will be provided during the presentation. Results for European Parliament elected in 2004 see in Turnovec [7].

Now let us calculate influence of regional party chapters in regionally coordinated voting, when all regional party chapters from the same region are voting together.

Region A: committee [11; 7, 10, 3]. Voting power of regional party chapters L, M, R in region A: (1/6, 4/6, 1/6). Total power of region A in the parliament regionally coordinated voting 1/6 is decomposed among the regional party chapters: (1/36, 4/36, 1/36)

Region B: committee [16; 15, 15, 0]. Voting power of regional party chapters L, M, R in region B: (1/2, 1/2, 0). Total power of region B in the parliament regionally coordinated voting 1/6 is decomposed among the regional party chapters: (3/36, 3/36, 0).

Region C: committee [26; 3, 22, 25]. Voting power of regional party chapters in region C: (1/6, 2/60, 3/6). Voting power of region C in the parliament regionally coordinated voting 4/6 is decomposed among the regional party chapters: (4/36, 8/36, 12/36)

Evaluation of voting power of regional party chapters in regionally coordinated voting is provided in Table 3. Entries in last column provide total voting power of each region in the parliament in the case of regionally coordinated voting, entries in the last row total voting power of each party in the parliament in the case of regionally coordinated voting, and entries in the main body of the table provide voting power of each regional party chapter.

regions	parties (voting power)			
	L	M	R	Total
A	1/36	4/36	1/36	6/36
B	3/36	3/36	0	6/36
C	4/36	8/36	12/36	24/36
total	8/36	15/36	13/36	1

Table 3 Decomposition of power in regionally coordinated voting

Let us assume that (based on empirical evidence) in average 3/4 of voting acts are ideologically coordinated and 1/4 of voting acts are regionally coordinated. Then, from the following matrix equation

$$\frac{3}{4}\begin{pmatrix} 0 & \frac{4}{36} & 0 \\ \frac{12}{36} & \frac{4}{36} & 0 \\ 0 & \frac{4}{36} & \frac{12}{36} \end{pmatrix} + \frac{1}{4}\begin{pmatrix} \frac{1}{36} & \frac{4}{36} & \frac{1}{36} \\ \frac{3}{36} & \frac{3}{36} & 0 \\ \frac{4}{36} & \frac{8}{36} & \frac{12}{36} \end{pmatrix} = \begin{pmatrix} \frac{1}{144} & \frac{16}{144} & \frac{1}{144} \\ \frac{39}{144} & \frac{15}{144} & 0 \\ \frac{4}{144} & \frac{20}{144} & \frac{48}{144} \end{pmatrix}$$

we obtain the mathematical expectation of voting power of regional party chapters, super-regional parties and regional representations under assumption that ideologically coordinated voting takes place with probability ¾ and regionally coordinated voting with probability 1/4 (see Table 4).[3]

Regions	parties (voting power)			
	L	M	R	Total
A	1/144	16/144	1/144	18/144
B	39/144	15/144	0	54/144
C	4/144	20/144	48/144	72/144
Total	44/144	51/144	49/144	1

Table 4 Mathematical expectation of voting power

[3] As was mentioned above, for the European Parliament these probabilities are estimated as 0.8 for ideologically coordinated voting and 0.2 for regionally coordinated voting, see Hix, Noury and Roland [1].

5 Concluding remarks

We tried to show that it is possible to evaluate not only the influence of European political parties as entities in ideologically driven voting and of national representations as entities in nationally driven voting, as it is usually done in analytical papers (Holler and Kellermann [2], Hosli [3], Nurmi [5]) but also the influence of national chapters of European political parties both in ideological and national voting and national influence in ideological voting, as well as the European political parties influence in national voting.

It was demonstrated that different dimensions of voting (ideological, national) lead to different levels of influence of the same national party group, European political party and national representation. The findings of our model analysis open the problem of strategic considerations, such as coalition formation, that can go across the existing structure, e.g. coalition of a country representation with some European political party, or preferring national coordination of different party groups of the same country to ideological coordination (this problem was opened with respect to Poland in Mercik, Turnovec, and Mazurkiewicz [4]). There is a broad area for extensions of presented model.

Acknowledgements

This research was supported by the Grant Agency of the Czech Republic, project No. 402/09/1066 "Political economy of voting behavior, rational voters' theory and models of strategic voting"..

References

[1] Hix, S., Noury, A. G., and Roland, G.: *Democratic Politics in the European Parliament.* Cambridge University Press, Cambridge, 2006.

[2] Holler, M., and Kellermann, J.: Power in the European Parliament: What Will Change? *Quantity and Quality*, **11** (1977), 189-192.

[3] Hosli, M..O.: Voting strength in the European Parliament: the influence of national and of partisan actors. *European Journal of Political Research* **31** (1977), 351-366.

[4] Mercik, J. W., Turnovec, F., and Mazurkiewicz, M.: Does voting over national dimension provide more national influence in the European Parliament than voting over ideological dimension. In: *Integration, Trade, Innovation and Finance: From Continental to Local Perspectives*, (Owsinski, J. W., ed.), Polish Operational and Systems Research Society, Warszawa, 2004, 173-186.

[5] Nurmi, H.: The representation of voters' groups in the European Parliament, a Penrose-Banzhaf index analysis. *Electoral Studies*, **16** (1977), 317-327.

[6] Shapley, L. S., and Shubik, M. A Method for evaluation the distribution of power in a committee system. *American Political Science Review* **48** (1954), 787-792.

[7] Turnovec, F.: Duality of power in the European Parliament. *IES Working Paper Series*, 6/2008.

Regional unemployment disparities and their dynamics: Evidence from the Czech Republic

Michal Tvrdoň[1], Tomáš Verner[2]

Abstract. Regional disparities have become a debated topic in the last two decades. Aim of this paper is to estimate trends in regional disparities at the level NUTS 3 during the period 2005-2011. This paper looks at unemployment in the Czech Republic and its persistence over time. The Czech labor market has faced many changes during this period, which were reflected not only at the national but also at the regional level. It also attempts to identify the factors responsible for this persistence. Regional unemployment trends during this period show that they have been characterized by a high degree of synchronicity, and regional unemployment disparities by a remarkable persistence. This persistence mostly reflects the existence of an underlying interregional equilibrium structure of unemployment differences, rather than prolonged inter-area labor market disequilibrium. Another aim of this paper is to test if regional and national unemployment rates are co-integrated, in other words if the long-term relationship between regional and national unemployment rates exists. Monthly Czech Ministry of Labor and Social Affairs data were used for the analysis. The Johansen test was applied on 2005-2011 data to examine co-integration between the regional and national unemployment rates. On the basis of the unit root test, we found that both variables are integrated of order I(1). Cointegration relationship thus has been demonstrated for Praha and the national unemployment rate only. So does the ECM. ECM implies deviation from equilibrium in short-run. About 23 % of deviation is corrected each month.

Keywords: cointegration, unemployment, regional disparities.

JEL Classification: J64, R11, R12

AMS Classification: 62P20

1 Introduction

Unemployment is a phenomenon in the economy when there is a mismatch between labor demand and labor supply. It means that greater labor supply quantity exists than is demanded. In other words, unemployment (or joblessness), as defined by the International Labour Organization, occurs when people are without jobs and they have actively sought work within the past four weeks. The unemployment rate is a measure of the prevalence of unemployment and it is calculated as a percentage by dividing the number of unemployed individuals by all individuals currently in the labor force.

Existing empirical studies have shown the correlation between the phases of the business cycle and changes in the unemployment rate. Typically, the number of available jobs is rising and the unemployment rate is declining in a time of economic growth. On contrary, the unemployment rate is increasing during the economic crisis. The question whether these changes take place with the same intensity at the regional level [1].

The aim of this paper is to estimate, based on econometric approach, relationship between the national rate of unemployment and regional rates of unemployment in the Czech Republic. In the parlance of economic time series analysis, this is equivalent to hypothesizing that regional and national unemployment rates are co-integrated, in other words that a stable linear relationship exists between these two variables. For this purpose the paper is divided into several parts. The first part is devoted to description of analytical tools. The second part contains empirical results of cointegration analysis and the final section summarizes all the key findings. We used Czech Statistical Office and Ministry of Labor and Social Affairs annual and monthly data between the year 2005 and 2011.

[1] Silesian University/School of Business Administration in Karviná, Department of Economics, Univerzitní nám. 1934/3, Karviná 733 40, tvrdon@opf.slu.cz.
[2] Silesian University/School of Business Administration in Karviná, Department of Economics, Univerzitní nám. 1934/3, Karviná 733 40, verner@opf.slu.cz.

2 Econometric methodology

Cointegration is an econometric technique for testing the relationship between non-stationary time series variables. This technique is often used because of many macroeconomic time series are not stationary in their levels. If two or more series each have a unit root, that is I(1), but a linear combination of them is stationary, I(0), then the series are said to be cointegrated. Thus cointegration analysis is an extension of the simple correlation based analysis. The objective of this article is to analyze the effects of economic growth on unemployment in the Visegrád group countries.

The problem then is to find a way to work with two possibly non-stationary series in a fashion that allows us to capture both short run and long run effects. In more technical parlance, cointegration is the link between integrated processes and steady state equilibrium. If the time series are stationary in first differences than it is fulfilled requirements for the implementation of cointegration. Although we have two non-stationary time series, their common cointegration long-term shift in time moves towards some equilibrium.

We used Phillips-Perron (PP) test as the unit root test. We used this approach to test the null hypothesis that a time series in integrated of order 1. The PP method estimates the non-augmented DF test equation, and modifies the *t*-ratio of the α coefficient so that serial correlation does not affect the asymptotic distribution of the test statistic. The PP test makes a non-parametric correction to the t-test statistic.

The PP test is based on the statistic [9]:

$$\bar{t}_\alpha = t_\alpha\left(\frac{\gamma_0}{f_0}\right)^{1/2} - \frac{T(f_0-\gamma_0)(se(\hat{\alpha}))}{2f_0^{1/2}s} \tag{1}$$

where $\hat{\alpha}$ is the estimate, and t_α the *t*-ratio of α, $se(\hat{\alpha})$ is coefficient standard error, and *s* is the standard error of the test regression. In addition, γ_0 is a consistent estimate of the error variance. The remaining term, f_0, is an estimator of the residual spectrum at frequency zero.

Cointegration test is based on the determination of r cointegration relations in the VAR model. Cointegration is confirmed, if true, that r > 0. For testing purposes, we used Johansen cointegration test [6].

It is necessary to obtain an indication of optimal time delay before the implementation of Johansen cointegration test, which was in our case according to the Schwarz information criterion (SC) applied to estimate the VAR model of differentiation two periods. The SC criterion is defined as [5]:

$$SC = n^{k/n}\frac{\sum \hat{u}^2}{n} = n^{k/n}\frac{RSS}{n} \tag{2}$$

where RSS means the residual sum of squares, k/n is the penalty factor.

We used two tests for determining the number of cointegration vectors: (i) the Trace test; and (ii) the Maximal Eigenvalue test.

The Trace test for the number of cointegrating vectors determines the number of cointegrating equations r:

$$(r) = -N^* \sum_{i=r+1}^{m} \ln(1-\hat{\lambda}_i) \tag{3}$$

We tested hypothesis by the Trace test for H_0 r=0 (there are no cointegration vectors) and H_1 r≤1 (there is cointegration equation). We did not reject the H_0 hypothesis if the Trace statistics is no larger than the 5% critical value).

Another test is the Maximal Eigenvalue test:

$$(r, r+1) = -N^* \ln(1-\hat{\lambda}_{r+1}) \tag{4}$$

We tested hypothesis by the Maximal Eigenvalue test for the same H_0 and H_1 like the Trace test. We also do not reject the H_0 hypothesis if the Maximal Eigenvalue statistics is no larger than the 5 % critical value.

As the following step the Error Correction Term (ECT) should be estimated and test for stationarity. The result of the PP test for the unit root should confirm integration in order I(0). It means that the Y_t and X_t are cointegrated or that the regression of equation in no longer spurious, and we can also find the linear combination

that connects Y_t and X_t in the long run [3] or we can say that there is a long-run equilibrium relationship between X and Y:

$$\hat{\mu}_t = Y_t - \hat{\beta}_1 + \hat{\beta}_2 X_t \tag{5}$$

Finally the Error Correction Model (ECM) should be estimated (if Y_t and X_t are cointagrated). Thus, we can express the relation between Y_t and X_t with an ECM specification as [3]:

$$\Delta Y_t = \alpha_0 - \alpha_1 (Y_{t-1} - \beta_1 X_{t-1}) + \beta_0 X_t + \varepsilon_t \tag{6}$$

where current changes in Ψ are a function of current changes in Ξ (the first difference of Ξ) and the degree to which the two series are outside of their equilibrium in the previous time period. Specifically, β_0 captures any immediate effect that Ξ has on Ψ, described as a contemporaneous effect or short-term effect. The coefficient, β_1 reflects the equilibrium effect of Ξ on Ψ. It is the causal effect that occurs over future time periods, often referred to as the long-term effect that Ξ has on Ψ. Finally, the long-term effect occurs at a rate dictated by the value of α_1.

Arlt [2] shows that it should be stressed that the importance of the ECM lies in the fact that it allows us to combine statistical and econometric approach to modelling economic time series.

3 Empirical results

Unemployment in the New Member States of the European Union has had both cyclical and structural nature. Moreover, the long-term lack of some professions existed in some regions. On the other hand, we can find some regions where a surplus of offered jobs existed. Another aspect, that prevents the matching process of supply and demand, is the lack of mobility and flexibility of the labor force. For this reason economic growth does not contribute significantly to improving labor market performance in the affected regions. Structural changes are continuing in some regions of the Czech economy, ie changes in the sectoral (industry) structure of the economy and the related changes in the professional and qualification structure of the labor force - during the transition process has been a marked shift of labor from the primary and secondary to the tertiary sector, while in the tertiary sector recorded the steepest rise in banking and insurance industry. Measured by the share of sectors in total employment in the economy the Czech economy belongs to the advanced economies of the world.

The main question is if this trend was reflected also in a regional context - in some regions were much more complex structural changes, mainly due to the current structure of the economy of these regions. The influence of these structural changes in different regions has different intensity, and among the most affected regions were the regions with large concentrations of heavy industry. The process of structural changes significantly influences the differences in unemployment rates among Czech NUTS III regions. The huge problem for maintaining the economy's growth rate, among other aspects, was a persistence of unemployment, which is significantly higher in the Czech Republic than in other EU countries. As stated by the Ministry of Labor and Social Affairs, the basic problems of regional unemployment is low labor mobility, qualifications of job seekers, particularly the long-term, which does not meet the labor market, a high proportion of job seekers with lower education and last but not least, it is the qualification structure of labor supply which does not correspond with needs of labor demand.

If we look at empirical data concerning unemployment in particular regions of the Czech Republic´s regions (see Table 1), we could assume that the lowest unemployment rate will be in metropolitan region. This assumption has proved to be valid; however it is possible to reproach some other implications, which are by their nature rather surprising. If differences between the unemployment rate reached in the metropolitan areas and the regions with the highest unemployment rate were significant (sustained period of high regional disparities in unemployment indicates low labor market flexibility mobility of the population, especially low regional mobility – [4]), similar relation for long-term unemployment was not so remarkable. Although the remarkable decrease of the unemployment rate has been recorded in the problematic regions Ústecký and Moravskoslezský kraj since the year 2005, the number of unemployed has stayed higher in these regions in comparison with other regions and it means a longstanding problem of highly regionalized structural unemployment. This is partly because of wide geographic diversity in a level of structural reforms and dynamics of economic growth, but also because of weak labor mobility. Two parallel phenomena occurred simultaneously in all Czech NUTS 3 regions – (i) a decrease of the total unemployment rate in all regions in the period 2004-2008; (ii) a significant increase of the share of long-term unemployment in total unemployment with its peak in 2006; and (iii) deterioration of labor market performance during 2009. However, intensity of the unemployment rate decline was quite different dur-

ing the observed period between 2004 and 2008. The unemployment rate did not decrease with the same intensity in Czech regions and we can state that its change ranged from –2.2 p.p. to -7.4 p.p. If we look at higher values of the unemployment rate in problematic regions Ústecký and Moravskoslezský kraj, we can assume persisting problems in these regions. Unsatisfactory labor market performance was confirmed by another indicator in these regions – the share of long-term unemployment in total unemployment. This share exceeded 60 % in some years, which means that six out of ten were unemployed for more than 12 months. Another finding is that this share was increasing gradually during the observed period, until outbreak of the economic crisis. An increase of the number of unemployed was one among consequences of the crisis and thus increasing the denominator in the formula for calculating the share of long-term unemployment, which resulted in a reduction of the share. Higher unemployment rate in these regions means also lower competitiveness (for more detailed analysis see [8]). The same trend was noticed on a national level. What is interesting is the fact that this trend was associated with all regions with no exceptions, even region Praha which still stayed below the whole national average. However, the share of long-term unemployment in total unemployment, which was over 39.2 % in 2008, is too high for the region with the highest concentration of foreign capital, a strong tertiary sector and the highest GDP per capita in the country. We take the view that this finding validates considerations that many of the unemployed are in principle unemployable in the Czech Republic due to the lavish social system and even though they meet conditions for inclusion into the category of unemployed, they are not its part de facto.

	2005	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**
Czech Republic	7.9	7.2	5.3	4.4	6.7	7.3	6.7
Praha	3.4	3.0	2.5	2.1	3.0	3.9	4.0
Středočeský kraj	6.3	5.7	4.6	4.0	5.8	7.1	7.1
Jihočeský kraj	6.3	6.0	4.8	4.0	6.5	7.4	7.3
Plzeňský kraj	6.4	5.9	4.9	4.2	7.0	7.8	7.1
Karlovarský kraj	10.2	9.5	8.0	6.9	9.9	10.8	10.2
Ústecký kraj	15.4	14.5	12.2	9.9	12.4	13.4	12.9
Liberecký kraj	7.8	7.4	6.5	6.0	10.0	10.6	9.6
Královéhradecký kraj	7.3	6.6	5.2	4.2	6.8	7.7	7.2
Pardubický kraj	8.3	7.3	5.8	5.0	8.0	9.1	8.3
Vysočina	8.2	7.4	6.1	5.2	8.7	9.6	9.1
Jihomoravský kraj	10.1	9.2	7.6	6.2	8.9	10.2	9.6
Olomoucký kraj	11.0	9.6	7.4	6.2	10.2	11.5	11.1
Zlínsky kraj	9.2	8.4	6.6	5.5	9.1	10.4	9.4
Moravskoslezský kraj	14.7	13.4	11.0	8.4	11.1	11.9	11.3

Table 1 Regional Unemployment Rates (annual data 2005-2011)

The data used for cointegration analysis were the national and regional unemployment rates between the years 2005 and 2011. We can find another ways how to estimate regional differences like DEA method (e.g. see Nevima and Ramik [10] or indices of geographical concentration (for more see Tvrdon [11]). We applied monthly data from the Czech Ministry of Labor and Social Affairs database. These data were not seasonally adjusted. We have selected four regions – two problematic regions with high unemployment (Moravskoslezský kraj and Ústecký kraj) rates and two regions with traditionally low unemployment rates.

We applied methodology which is taken from Martin [7]. Ústecký and Moravskoslezský region have the greatest average unemployment rate; on the contrary Praha and Středočeský region have the smallest average unemployment rate (see Figure 1). We used monthly data between January 2005 and December 2011. We found out time series are integrated of order 1, I(1), i.e. stationary at the first difference.

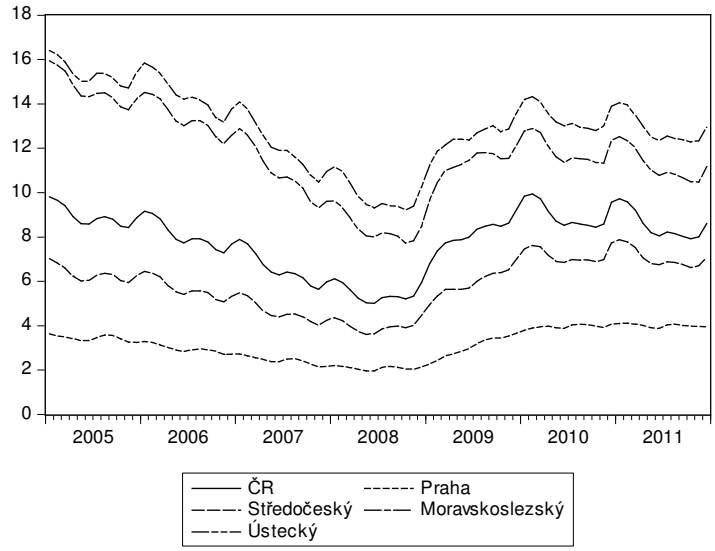


Figure 1 Unemployment rate, monthly data (%)

Next step was to obtain an indication of optimal lag length before performing Trace test and Maximum Eigenvalue test. The optimal time delay through Schwarz information criterion equals to 12. Trace test and Maximum Eigenvalue were performed to determine the number of *r* and ECM. Results of the unrestricted cointegration test can be seen in Table 2[3].

Area	H_0	Trace statistic	5 % critical value	Probability 5 %
Praha	r = 0	48.65562	20.26184	0.0
	r = 1	3.948413	9.164546	0.4199
Středočeský	r = 0	44.00313	20.26184	0.0
	r = 1	4.528271	9.164546	0.3391
Moravskoslezský	r = 0	53.01105	20.26184	0.0
	r = 1	5.954942	9.164546	0.1942
Ústecký	r = 0	60.15372	20.26184	0.0
	r = 1	7.211175	9.164546	0.1157

Table 2 Unrestricted cointegration rank test (Trace test)

The null and alternative hypotheses were established to determine the number of *r*: H_0: r = 0 against H_1: r > 0. We reject the null hypothesis at 5 % significance level; therefore the number of cointegrating equations is not equal to zero. In next round we established: H_0: r = 1 against H_1: r > 1. We accepted null hypothesis, thus the number of cointegrating equation *r* equals to one.

Based on previous tests of cointegration, we yield the following equations:

$$CR = 1.170206\,PRA + 4.381402 + e_t \tag{7}$$

$$CR = 0.494633\,STR + 5.099111 + e_t \tag{8}$$

$$CR = -0.099610\,MOR + 9.785900 + e_t \tag{9}$$

$$CR = -0.198305\,UST + 11.10721 + e_t \tag{10}$$

where CR, PRA, STR, MOR, UST denotes unemployment rates of the Czech Republic, Středečeský, Moravskoslezský and Ústecký region. If we look at equations (9) and (10) we can see negative relationship which is not consistent with economic theory and variables do not trend together.

Time series of errors (ECT) were non-stationary at levels, therefore they are not integrated of order 0, I(0); except relationship between unemployment rates of the Czech Republic and Praha region thus is confirmed for

[3] We present the trace results only for economy of space.

these time series only. Increasing unemployment in Praha about 1 percentage point will cause increasing unemployment rate almost 1.2 percentage point in the Czech Republic. ECM implies deviation from equilibrium in short-run. About 23 % of deviation is corrected each month.

4 Conclusion

The aim this paper was to perform an empirical analysis of long-term relationship between national and regional unemployment rates of selected regions. In other words, we try to find possible cointegration between changes of the national unemployment rate and the regional unemployment rates. On the basis of the unit root test, we found that in all regions, unemployment rates are integrated of order I(1). This result allowed us to continue and after establishing a period of lag (twelve periods), we implemented the Johansen cointegration test. This test showed that although one cointegration relationship. ECT, however, showed that residues are not stationary at I(0) except relationship between Prague and national unemployment rates. Cointegration relationship thus has been demonstrated for Praha and the national unemployment rate only. So does the ECM. ECM implies deviation from equilibrium in short-run. About 23 % of deviation is corrected each month.

Acknowledgements

This research was financially supported by the Student Grant Competition SU within the project SGS/7/2012 "The influence of regional disparities on the business environment".

References

[1] Abraham, KG., and Shimer, R.: *Changes in Unemployment Duration and Labor Force Attachment*. NBER Working Paper No. 8513. NBER. Cambridge (MA). 2001.
[2] Arlt, J.: Kointegrace v jednorovnicových modelech. *Politická ekonomie* **45** (1997). 733-746.
[3] Asteriou, D., and Hall S.G.: *Applied Econometrics*. Palgrave MacMillan. New York. 2007.
[4] Czech National Bank, Analyses of the Czech Republic´s current economic alignment with the euro area, Prague: CNB, 2006.
[5] Gujarati, D., and Porter. DC.: *Basic Econometrics*. 5th ed. McGraw-Hill. Boston. 2009.
[6] Johansen, S.: *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford University Press, 1995.
[7] Martin, R.: Regional Unemployment Disparities and their Dynamics. Regional Studies **31** (1997). 237-252.
[8] Muntean, MC., Nistor, R., and Nistor, C.: Competitiveness of Developing Regions in Romania. WSEAS Transactions on Business and Economics **7** (2010). 252-261.
[9] Phillips. P. and Perron. P.: Testing for a Unit Root in Time Series Regression. *Biometrika* **75** (1988). 335-346.
[10] Nevima, J., and Ramik J.: Application of DEA for evaluation of regional efficiency of EU regions. In Proceedings of the 28th International Conference on Mathematical Methods in Economics 2010. České Budějovice: University of South Bohemia, Vol. I and II, pp. 477-482.
[11] Tvrdon, M.: Regional Income Disparities in the New Member States: De-composition and Measurement. In Proceedings of the 28th International Conference on Mathematical Methods in Economics 2010, České Budějovice: University of South Bohemia, Vol. I and II, pp. 640-645.

Financial accelerator mechanism in a small open economy: DSGE model of the Czech economy

Stanislav Tvrz[1], Jaromír Tonner[2], Osvald Vašíček[3]

Abstract. Our research is motivated by current experience of worldwide economic difficulties that suggests that it could be the banking sector that can to a certain extent cause and exacerbate economic recessions. In this paper we focus on the effects of financial accelerator mechanism in the Czech economy in recent period of unstable financial sector and the debt crisis in the EU. In order to be able to assess the importance of the financial frictions in this particular economy, a DSGE model of a small open economy is estimated. Model framework containing financial accelerator mechanism proposed by Bernanke, Gertler and Gilchrist (1999) is used for the analysis. The original model specification is slightly altered by adding a shock in entrepreneurial net-worth. Also, we decided to model the foreign sector as a VAR block, which enables us to impose more structure on foreign variables than independent AR processes. Quarterly data from period 1999 to 2011 are used for the estimation. Model parameters are estimated with the use of Bayesian techniques. A method of shock decomposition is used to analyze historical development of endogenous variables and to evaluate particular effects of individual exogenous shocks.

Keywords: DSGE model, financial frictions, financial accelerator, Bayesian methods, variance decomposition, shock decomposition.

JEL classification: E44, E32
AMS classification: 91B64

1 Introduction

In the aftermath of the global financial crisis of 2007–2009 that provoked subsequent global economic slowdown and recently in some European countries even a debt crisis, the economists around the world realize that there exist important relations between the financial sector and the real economy. In fact, current experience of worldwide economic difficulties shows that the macro-financial linkages can to a certain extent influence the aggregate fluctuations.

Thanks to high capitalization, liquidity and rentability of the Czech banking system, the Czech economy was exposed to the subprime crises only marginally. Nevertheless, the global economic economic crisis hit the Czech economy in the last quarter of 2008. The development of the domestic economy during the global crisis was influenced predominantly by the downturn of foreign demand.

Despite the fact that the impacts of the crisis on the domestic financial sector were relatively mild, the interest rate spreads increased substantially. During the period of economic boom of 2006–2008 the interest rate spreads declined slightly. With the outburst of the global economic crisis the difference between the commercial interest rates and the policy interest rate increased sharply, which significantly reduced the efficiency of monetary policy in this period. The central bank started lowering the policy interest rate in 2009. However, much of its effects were countered by the growing spread.

Apparently, there can be times when the development of the policy interest rate differ significantly from the development of commercial interest rates that the households and firms in the economy actually face. The idea of a frictionless banking sector seems, therefore, no longer tenable. Several financial

[1]Masaryk University, Department of Economics, Lipová 41a, 602 00 Brno, olorin@mail.muni.cz
[2]Masaryk University, Department of Economics, Lipová 41a, 602 00 Brno and Macroeconomic Forecasting Division, Czech National Bank, Na Příkopě 28, 115 03 Praha, Jaromir.Tonner@cnb.cz
[3]Masaryk University, Department of Economics, Lipová 41a, 602 00 Brno, osvald@mail.muni.cz
The views expressed here are those of the authors, and do not necessarily reflect the position of the Czech National Bank.

frictions mechanisms were considered in academia as well as in the central banks in past years that would enhance current workhorse DSGE macroeconomic models with mechanisms that could explain what we observe. In our paper we use the model framework developed by Shaari [6] that includes financial accelerator mechanism proposed by Bernanke *et al.* [2]. We introduce a net worth shock into the original model and redefine the foreign sector as a VAR(1) block. 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. Using the estimated DSGE model, we employ a method of shock decomposition in this paper to analyse historical development of real output and to evaluate particular effects of individual exogenous shocks with special attention given to the shocks in the entrepreneurial net worth.

2 The model

The model framework follows Shaari [6]. Overall structure of the model is based on Galí and Monacelli [4] and it is modified to incorporate the financial accelerator mechanism in line with Bernanke *et al.* [2].

The model contains households, entrepreneurs, retailers, central bank and foreign sector. 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 definition of debt-elastic risk premium follows Adolfson *et al.* [1] and it contains exogenous AR(1) component of risk-premium or uncovered interest parity shock (A_t^{UIP}). The households then spend their earnings on consumption and domestic and foreign bonds acquisition.

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 (A_t^Y) and the capital goods production is subject to capital adjustment costs. Entrepreneurs finance the production and ownership of capital by their net-worth N_t and borrowed funds F_t. The 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 real price of capital or Tobin's Q and χ is financial accelerator parameter. To maximize profit the entrepreneurs choose the level of capital K_{t+1} and the level of borrowed funds F_{t+1} in accordance with following optimality condition,

$$E_t(R_{K,t+1}) = E_t\left[\left(\frac{N_{t+1}}{Q_t K_{t+1}}\right)^{-\chi} R_t \frac{P_t}{P_{t+1}}\right], \tag{2}$$

where $R_{K,t+1}$ is marginal return from capital investment, R_t is nominal interest rate, P_t is price level and E_t is expectations operator. The entrepreneurial equity develops according to

$$V_t = \left[R_{K,t}Q_{t-1}K_t - \left(\frac{N_t}{Q_{t-1}K_t}\right)^{-\chi} R_{t-1}\frac{P_{t-1}}{P_t}F_t\right]. \tag{3}$$

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 transfered 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 effective survival rate of entrepreneurs. Its logarithmic deviation from steady state is assumed to evolve according to AR(1) process. Entrepreneurs also receive wage $W_{E,t}$ for the labour they supply to the production of domestic intermediate goods. Entrepreneurial net-worth is then given by

$$N_{t+1} = \varsigma A_t^{NW} V_t + W_{E,t}. \tag{4}$$

Next, there are two types of retailers in the model. Home goods retailers and foreign goods retailers. Both types of retailers are assumed to operate in conditions of monopolistic competition. Home good

retailers buy domestic intermediate goods at wholesale price and sell the final home goods to the consumers. Foreign good retailers buy goods from foreign producers at the wholesale price and resell the 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 (A_t^{LOP}). By Calvo-type price setting and inflation indexation of the retailers the nominal rigidities are introduced into the model.

The central bank determines the nominal interest rate in accordance with forward-looking Taylor type interest rate rule. Deviations of interest rate from the interest rate rule are explained as monetary policy iid shocks (ε_t^{MP}).

Following Christiano *et al.* [3], the foreign economy variables - real output, CPI inflation and nominal interest rate, are modelled using a VAR(1) model of this form,

$$\begin{pmatrix} y_t^* \\ \pi_t^* \\ r_t^* \end{pmatrix} = \begin{pmatrix} \rho_{y^*y^*} & \rho_{y^*\pi^*} & \rho_{y^*r^*} \\ \rho_{\pi^*y^*} & \rho_{\pi^*\pi^*} & \rho_{\pi^*r^*} \\ \rho_{r^*y^*} & \rho_{r^*\pi^*} & \rho_{r^*r^*} \end{pmatrix} \begin{pmatrix} y_{t-1}^* \\ \pi_{t-1}^* \\ r_{t-1}^* \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ \sigma_{\pi^*y^*} & 1 & 0 \\ \sigma_{r^*y^*} & \sigma_{r^*\pi^*} & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_t^{y^*} \\ \varepsilon_t^{\pi^*} \\ \varepsilon_t^{r^*} \end{pmatrix}, \tag{5}$$

where $\varepsilon_t^{y^*} \sim iid(0, \sigma_{y^*}^2)$, $\varepsilon_t^{\pi^*} \sim iid(0, \sigma_{\pi^*}^2)$ and $\varepsilon_t^{r^*} \sim iid(0, \sigma_{r^*}^2)$. Autocorrelation coefficients satisfy $\rho_{y^*y^*}, \rho_{\pi^*\pi^*}, \rho_{r^*r^*} \in \langle 0, 1 \rangle$. Remaining coefficients are not constrained. Compared to foreign variables modelled as independent AR(1) processes this approach should capture the relations between foreign variables more closely.

3 Estimation

The 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 75 % of draws were discarded as burn-in sample. The scale parameter was set to achieve the acceptance rates around 30 %.

3.1 Data

Quarterly time series of eight observables were used for the purposes of estimation. These time series cover the period between the first quarter of 1999 and the fourth quarter of 2011 and contain 52 observations each.

For the domestic economy, time series of real aggregate product, consumer price index (CPI), 3-month PRIBOR and Prague stock exchange (PSE) PX index are used. These seasonally adjusted time series were obtained from the database of Czech Statistical Office and PSE.

The foreign economy is represented by eleven founding countries of the Euro area plus Greece that joined the monetary union in 2001. For these twelve economies, the seasonally adjusted time series of real aggregate product, CPI index and 3-month EURIBOR are used. Also, time series of CZK/EUR real exchange rate is used. These data were obtained from the Eurostat database.

The time series of PX index is used as a proxy for the entrepreneurial net worth. Changes in the value of the companies traded on the PSE are expressed in a condensed form in the PX index. We assume that all the companies in the economy are affected by the same events as the companies traded on the PSE and that they are affected in a similar way. Therefore, we come to a conclusion that the development of the PX index should to a reasonable extent capture also the changes in the aggregate value of all the companies in the economy and therefore also the aggregate entrepreneurial net worth.

The original time series were transformed prior to estimation so as to express the logarithmic deviations from steady state. Log values of the CPI indices were taken and their first differences were calculated. Obtained time series were demeaned and identified as the logarithmic deviations of the domestic and foreign CPI inflation from steady state. The logarithmic deviations of remaining observables from their steady state were calculated with the use of Hodrick-Prescott filter (HP filter). Since we are working with quarterly time series in this paper, HP filter with parameter $\lambda = 1600$ was used to find the approximation of their steady state.

3.2 Priors and posteriors

The DSGE model was successfully estimated on historical data of the Czech economy in the period 1999–2011. Prior and posterior distributions of the structural parameters are described in table 1. Obtained estimates are similar to the results reported in the literature.

The parameter of the financial accelerator is estimated at a value of 0.0379, which is a slightly higher value that the estimate of 0.0269 reported by Tonner and Vašíček [7]. The parameter Γ is estimated at the value of 1.8407, which is below the prior mean of 2. Our result is quite close to Bernanke *et al.* [2] who suggest calibration of this parameter to 2. The entrepreneurs' survival rate ς is estimated somewhat below the prior mean at the value of 0.9332, which implies average entrepreneurs' business lifespan of nearly 4 years. Tonner and Vašíček [7] calibrate this parameter to the value of 0.9728, which would imply the business lifespan of almost 10 years.

Interest rate smoothing parameter of the Taylor rule is estimated at the value of 0.6379. Similar value of 0.6647 is reported by Tonner and Vašíček [7]. The posterior mean of parameter β_π suggests a value of inflation weight in the Taylor rule of 1.7260. This result is in accordance with Ryšánek *et al.* [5] who found a value of 1.68. The weight of output gap in the Taylor rule Θ_y is estimated at 0.5772, which is close to the result of 0.5288 reported by Tonner and Vašíček [7].

			Prior		Posterior		
Parameter		**Distribution**	**Mean**	**Std**	**Mean**	**95% HPDI**	
Υ	habit persistence	Beta	0.60	0.05	0.5991	0.5112	0.6833
Ψ	inverse elast. of labour supply	Gamma	2.00	0.50	1.7852	1.0604	2.4561
ψ^B	elast. of debt-elastic risk prem.	Gamma	0.05	0.02	0.0393	0.0182	0.0612
η	home/foreign goods elast. subst.	Gamma	0.50	0.10	0.4885	0.3825	0.5956
κ	price indexation	Beta	0.50	0.10	0.2881	0.1705	0.4053
γ	preference bias to foreign goods	Beta	0.65	0.10	0.5236	0.4260	0.6199
θ_H	home goods Calvo parameter	Beta	0.75	0.10	0.7494	0.6913	0.8108
θ_F	foreign goods Calvo paremeter	Beta	0.75	0.10	0.7913	0.7341	0.8498
ψ^I	capital adjustment costs	Gamma	20.0	5.00	26.085	17.059	34.369
Financial frictions							
Γ	s-s. capital/net worth ratio	Gamma*	2.00	0.50	1.8407	1.3360	2.3197
ς	entrepreneurs' survival rate	Beta	0.973	0.015	0.9332	0.8988	0.9684
χ	financial accelerator	Gamma	0.05	0.02	0.0379	0.0133	0.0621
Taylor rule							
ρ	interest rate smoothing	Beta	0.70	0.10	0.6379	0.5643	0.7169
β_π	inflation weight	Gamma*	1.50	0.2	1.7260	1.3713	2.0669
Θ_y	output gap weight	Gamma	0.50	0.2	0.5772	0.3799	0.7775

*Shifted gamma distribution is used for these parameters, because they are assumed to take values from interval $\langle 1, \infty)$.

Table 1 Estimated structural parameters

3.3 Variance decomposition

The table 2 contains asymptotic variance decomposition of model variables. The variance of the model variables is decomposed into the contributions of exogenous shocks based on a simulation with infinite horizon.

Obviously, the net worth shock is a one of the driving forces of this model in the long run. More than half of the variance of real output, consumption, nominal interest rate and inflation is explained by this shock. Nearly 75 % of the variance of the real return to capital investment and over 85 % of the real investment, real price of capital and capital stock is explained by the net worth shock. The variance of entrepreneurial net worth and external finance premium is explained predominantly by this shock.

In general, we can say that the shocks in entrepreneurial net worth, domestic productivity, law of one price gap and foreign output explain predominant part of the variance of the model variables. The remaining shocks play only marginal role in the long run.

Variable		NW	Y	MP	UIP	LOP	r^*	π^*	y^*
y	real output	59.86	1.93	8.14	9.47	12.84	3.10	0.93	3.73
c	real consumption	59.34	1.66	1.56	6.79	20.52	0.47	0.77	8.89
inv	real investment	86.93	2.92	1.01	1.67	5.71	0.02	0.07	1.67
k	capital stock	87.29	2.98	0.40	0.26	7.29	0.00	0.04	1.73
r	nominal interest rate	63.50	2.63	1.52	11.27	13.60	0.53	0.36	6.59
rer	real exchange rate	26.17	0.76	1.45	4.31	64.08	0.84	0.33	2.05
π	CPI inflation	60.11	19.39	2.83	5.43	8.97	0.70	0.11	2.46
z	net foreign assets position	6.48	0.15	1.47	5.29	53.99	0.38	2.14	30.10
l_H	households' labour supply	1.82	83.52	3.35	4.19	3.89	1.35	0.41	1.47
w_H	households' wages	6.75	74.12	3.79	1.50	9.82	0.86	0.52	2.64
w_E	entrepreneurs' wages	2.33	81.20	3.73	2.10	7.21	1.05	0.47	1.92
r_G	gross rental rate of capital	17.36	69.27	3.23	1.81	5.48	0.89	0.39	1.56
r_K	gross return on capital invest.	72.16	16.44	7.72	1.28	1.81	0.05	0.12	0.43
q	real price of capital	86.83	2.90	1.16	2.03	5.31	0.03	0.08	1.66
mc	marginal costs	0.07	74.56	3.64	3.05	16.57	1.26	0.41	0.45
n	entrepreneurial net worth	95.17	3.35	0.65	0.09	0.54	0.03	0.01	0.16
efp	external finance premium	95.55	3.41	0.65	0.12	0.14	0.05	0.01	0.07
r^*	foreign interest rate	0.00	0.00	0.00	0.00	0.00	14.67	5.31	80.01
π^*	foreign CPI inflation	0.00	0.00	0.00	0.00	0.00	0.16	79.22	20.63
y^*	foreign output	0.00	0.00	0.00	0.00	0.00	0.58	4.70	94.72

Table 2 Variance decomposition (in percent)

3.4 Shock decomposition

In the following section, we discuss the shock decomposition of the real output gap, that is depicted in figure 1 and is expressed in percent of the potential output. The instrument of shock decomposition allows us to see the effects of particular exogenous shocks on this smoothed variable in time.

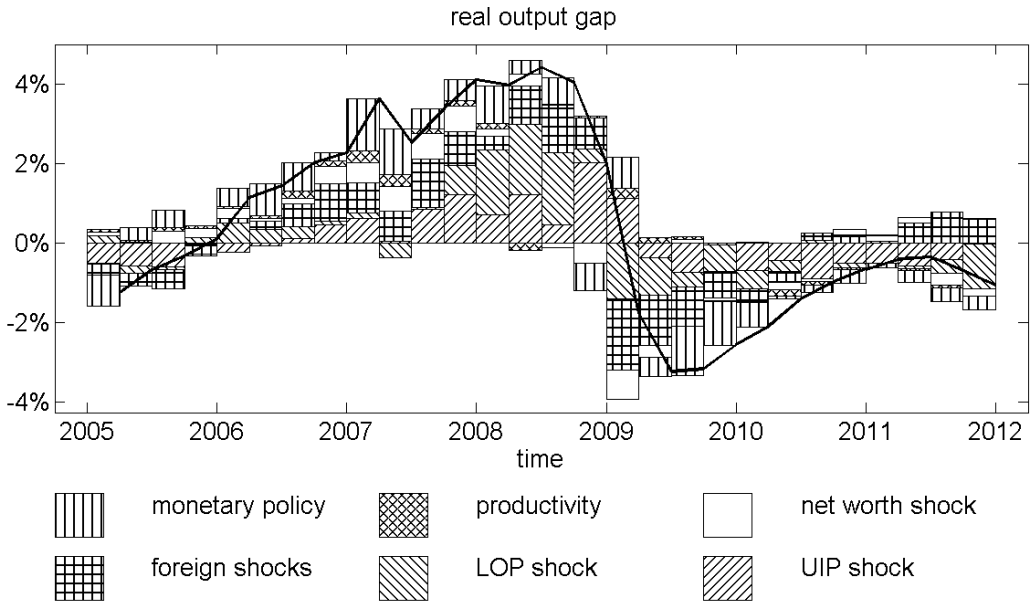


Figure 1 Shock decomposition of real output gap

According to the model, the foreign output shock played important role in the economic upswing of 2006–2008. Growing foreign output in this period boosted the foreign demand and net exports. Positive shock in the entrepreneurial net worth played its role as well, especially in 2007. Rather low nominal interest rate in this period also supported the economic growth. In the second half of 2007, the debt-elastic risk premium shock and the shock in the law of one price gap further stimulated the growth of real output.

In the last quarter of 2008, the foreign demand started to subside. Together with the exogenous decline in entrepreneurial net worth and relatively high nominal interest rate this caused a swift decline of the real output in this period. In the beginning of 2009, the shock in the law of one price gap caused a quick depreciation of the real exchange rate and further accelerated the decline of the real output. Low foreign demand affected the domestic real output mainly in the 2009 and in the beginning of 2010. The exogenous reduction of the entrepreneurial net worth took the biggest effect in the beginning of 2009, when it contributed around 0.75 percentage points to the negative real output gap. During 2010, the negative effects of the majority of exogenous shocks diminished and the real output returned to the neighbourhood of its potential.

4 Conclusion

It was a goal of this paper to evaluate the importance of financial frictions in the Czech economy during the period of global financial and economic crisis. A DSGE model framework proposed by Shaari [6] was chosen for this purpose. This model of a small open economy contains the financial accelerator mechanism introduced by Bernanke *et al.* [2] and it also includes important real and nominal rigidities. Original model was slightly modified by introducing exogenous shock in the entrepreneurial net worth into the model. Furthermore, the foreign sector was redefined as a VAR(1) block.

According to the variance decomposition, the shock in entrepreneurial net worth is one of the driving forces of the model economy. Together with the shock in domestic productivity, law of one price gap and the foreign output, it explains predominant part of the variance in the long run. The shock decomposition of the real output showed considerable effects of the shock in the net worth during the period of economic boom of 2006–2008 and also in the early phase of the economic crisis. Even though the development of the Czech economy in the period 2006–2010 was mainly affected by the situation abroad and especially by the development of foreign demand, the results of this model suggest relatively significant effects of the financial frictions as well.

Acknowledgements

This work is supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/0780/2011.

References

[1] Adolfson, M., Laséen, S., and Villani, M.: *Bayesian Estimation of an Open Economy DSGE with Incomplete Pass-Through.* Journal of International Economics, vol. 72, 2008, 481–511.

[2] Bernanke, B.S., Gerlter, M., and Gilchrist, S.: *The Financial Accelerator in a Quantitative Business Cycle Framework.* In: Taylor, J.B., Woodford, M (eds.) Handbook of Macroeconomics, vol. 1, Elsevier, Amsterdam, 1999, 1341–1393.

[3] Christiano, L. J., Trabandt, M., and Walentin, K.: *Introducing financial frictions and unemployment into a small open economy model.* Journal of Economic Dynamics and Control, vol. 35, Issue 12 (December), 2011, 1999–2041.

[4] Galí, J., and Monacelli, T.: *Optimal Monetary and Fiscal Policy in a Currency Union.* Economics Working Papers 909. Universitat Pompeu Fabra: Department of Economics and Business, 2005.

[5] Ryšánek, J., Tonner, J., and Vašíček, O.: *Monetary Policy Implications of Financial Frictions in the Czech Republic.* CNB Working paper series, Czech national bank, Prague, 2011.

[6] 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.

[7] Tonner, J., and Vašíček, O.: *Financial Frictions Relevance during the Crisis: Czech Case.* In: Advances in Applied Economics, Business and Development, vol. 209, Springer, Dalian, 2011, 339–344.

Analysis of the impact of selected variables on the availability of accommodation facilities

Klára Václavinková [1], Milena Botlíková [2], Miroslava Kostková [3]

Abstract. One of the evaluation criteria of success of tourism destinations are the outputs of accommodation facilities for tourism, it is the availability of accommodation facilities. Statistical data helps assess not only the domestic availability, as well as foreign participants. The subject of investigation will be the evaluation criteria in terms of occupancy residents and not residents. The actual criterion is influenced by many factors, such as substitution, crime, the amount of disposable income of the population of the CZ. etc. However, these effects have an impact not only on the level of demand of the accommodation, but also affect the selection of categories (hotels, cottages, campsites). Organizations providing services in the tourism sector, especially housing, creates incentives to accelerate the development of other sectors (trade, transport infrastructure, construction and others). These indicators are very sensitive to economic fluctuations. Any analysis will allow for the cross-sectional statistical data to establish relationships and dependencies of these quantities. The result will be the definition of needs for further development and the ability to create tools for improving these services.

Keywords: tourism, accommodation, services, attendance.

JEL Classification: C44, C01, C51
AMS Classification: 62P20, 62J12, 91B70

1 Introduction

In the last years, the development of tourism. Tourism is the main sectors involved in the economic development of the country, but it also has meaning. Not only is an important component of foreign trade in recent years to create a positive balance, creating working conditions not only in the travel industry, but also associated industries. The development of tourism is needed lately to pay attention to tourism through the implemented residents (It is estimated that with jobs in tourism tied 6-8 jobs tied to tourism). Tourism development has since 2004 fluctuating trend

Negative effects caused in the tourism market growth in supply over demand, which creates pressures to reduce prices, which may be beneficial for many companies liquidation. To maintain and attract a customer tries to individual facilities creating different types of promotions from various pricing advantage, service packages and other incentives. The problem is that the menu structure and targeting specific audiences is correct.

One of the possible causes of the decline in tourism can be caused due to:

- safety in the locality,
- environmental load, disaster,
- economic situation, price,
- legislative restrictions,
- standards of service.

Criteria for assessment of the situation in the tourism sector are the numbers of each occupancy accommodation or overnight stays. Accommodation facilities in the CZ since 2010 are classified according to a common European system, which integrates the services provided by businesses, facilities, and other criteria. Individual property is divided into hotels 1 to 4 three star hotels, motels and guest houses, camps and other cottage. European system classification works under the auspices of the Union Hotel Stars, owned by the Association of Hotels and Restaurants, hotels have classified this way to gain a competitive advantage such as promotion in all coun-

[1] Silesian University in Opava, department of Tourism Univerzitní nám. 1934/3, 733 40 Karviná, Czech Republic, vaclavinkova@opf.slu.cz Silesian University in Opava, School of Business Administration in Karvina, department of Logistics, Univerzitní nám. 1934/3, 733 40 Karviná, Czech Republic, botlikova@opf.slu.cz.
[2] Silesian University in Opava, department of Logistics, Univerzitní nám. 1934/3, 733 40 Karviná, Czech Republic, botlikova@opf.slu.cz.
[3] Silesian University in Opava, department of Tourism Univerzitní nám. 1934/3, 733 40 Karviná, Czech Republic, kostkova@opf.slu.cz.

tries registered in the system. Categorization for statistical surveys of the CSO then defines Hotels 4 and 3 star hotels, guesthouses, campsites, cottages and cottage settlements and others.

Current situation in tourism is associated with the crisis period, although the crisis was the cause of the collapse in demand for accommodation in 2008.2009, fluctuating trends have been observed since 2004, as evidenced by changes in the number of residents in each guest accommodation. If we came out of the above, then, if safety and environmental load can not be assumed reasons for reducing the demand for accommodation. CZ does not deal with other countries such as riots, strikes and is not associated with any serious environmental or natural disasters. The problem might seem economical and social situation of residents and standards of service.

The assumption is that domestic tourism will be bound to the exchange rate is correct. Number of residents in 2006 and 2007 the upward trend. Exchange rates at the time declined to strengthen the crown.

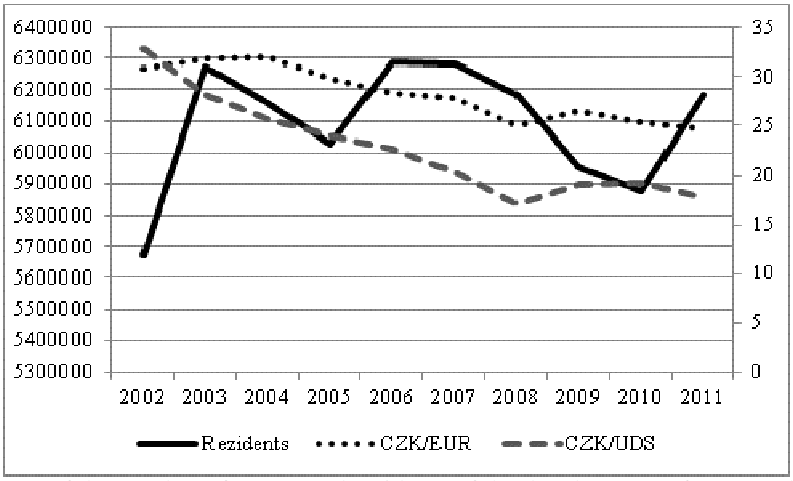


Figure 1 Development of the number of guests and residents of the development of UDS and the euro exchange rate [1]

1.1 Factors affecting interest in tourism

Evaluate the success of tourism is possible in the number of guests and overnight stays in different categories. Our goal is to determine what factors act on the selection of different types of accommodation. As described above for the decline in tourism can affect economic and social situation of residents exchange rates of CZK against the euro or dollar, or safety

In determining the occupancy dependence of individual accommodation facilities and exchange rate value of each course were recalculated on the basis of the values observed rate of the Czech National Bank (as the value of 2 consecutive, excluding weekend days) and converted to values of the quarter or annual averages, so that they are compatible with value for the number of guests for the period, overnight. The values of CZ and unemployment data on safety, based on the values of the Czech Statistical Office (CSO) [1].

1.2 Finding a suitable model

The evaluation will be based variable number of variables in each guest accommodation establishments regressors individual factors that are, according to the above-mentioned possible causes of the decline in tourism:

- the average wage,
- unemployment,
- rate CZK / EUR,
- CZK / USD exchange.

We use multiple regression model to determine the dependence between the variables.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + ...$$

The dependence between the number of guests in each accommodation and individual factors.
The regression model (1) was determined on the basis of explored values of guests of "*****" hotels and salary:

$$HOTEL\ 5* = -100\ 251 + 26.25 * pay \qquad (1)$$

The following values indicate that the regression model is chosen correctly. The coefficient of determination $R^2 = 0.51$, which indicates what proportion of the variance in the dependent variable observations. Regression was able to explain. Use the R^2 measure described tightness depending regression functions. The larger the value of R^2, ie value approaching to 1, the mean regression greater success and vice versa [5]. At the same time the value of the parameter $\beta 1 = 2.44\ E\text{-}07$ shows the significance factor wages (see Table1).

RESULT						
Regression statistics						
Correlation coefficient R	0.712931					
The coefficient determination R^2	**0.508271**					
Adjusting coefficient deternation R^2	0.495331					
Error mean	86072.343332					
Observation	40					
ANOVA						
	Unlike	*SS*	*MS*	*F*	*Significance F*	
Regresion	1	2.9099E+11	2.9099E+11	39.278307	**2.44893E-07**	
Residues	38	2.8152E+11	7,4084E+09			
Total	39	5,7251E+11				
	Coefficients	*Error mean*	*t Stat*	*p Value*	*Low 95%*	*Up 95%*
Limit	-100250.749	85822.45908	-1.1681179	0.250035984	-273989.23324	73487.734375
pay	26.2540538	4.189093	6.267241	**2.44893E-07**	17.773679	34.734429

Table 1 MS Excel regression analysis

For better explanation of the regression model will be examined in relationship to the unemployment figures. We assume that the unemployed persons, the accommodation of disinterest. The reason for this postponement is high and unnecessary financial costs. After inclusion of the variable unemployment there was a small increase in $R^2 = 0.508617$, $\beta_1 = 0.872609$ Parameter is insignificant, so it disposes of the variables.

Another variable regression model for improvement is to use the exchange rate CZK / EUR. We assume that both residents and nonresidents will move according to prefer accommodation at home or abroad. Again, there was only a slight improvement in the coefficient of determination $R^2 = 0.50965$, for reasons of immateriality parameter $\beta^1 = 0.749225$ is the elimination of variable exchange rate CZK / EUR.

Also, the variable exchange rate CZK / USD showed little growth and the insignificance of the parameter β_1, you see Table 2.

RESULT						
Regression statistics						
Correlation coefficient R	0.735326					
The coefficient determination R^2	**0.540704**					
Adjusting coefficient deternation R^2	0.515877					
Error mean	84302.0333					
Observation	40					
ANOVA						
	Unlike	*SS*	*MS*	*F*	*Significance F*	
Regresion	2	3.0956E+11	1.5478E+11	21.779007	**5.6069E-07**	
Residues	37	2.6295E+11	7106832822			
Total	39	5.7251E+11				
	Coefficients	*Error mean*	*t Stat*	*p Value*	*Low 95%*	*Up 95%*
Limit	390433.803709	314990.259151	1.239511	0.222961	-247797.080378	1028664.687796
pay	12.938552	9.202997	1.405906	**0.168095**	-5.708491	31.585595
CZK/UDS	-9776.229147	6048.174046	-1.616393	**0.114505**	-22030.993718	2478.535425

Table 2 MS Excel regression analysis

The final model of the relationship between the number of guests and the average wage in the CZ expresses the model (1).

In the evaluation of links between the number of guests in 4-star hotels and various factors was followed by analogy. In the first phase was interviewed by the dependence between the number of guests in the hotel paid. The regression model was created, the results of regression analysis, see Table 3:

$$HOTEL\ 4* = -1\,809\,144.35 + 176.69 * pay \qquad (2)$$

RESULT						
Regression statistics						
Correlation coefficient R	0.848870					
The coefficient determination R^2	**0.720581**					
Adjusting coefficient deternation R^2	0.713228					
Error mean	366735.051690					
Observation	40					
ANOVA						
	Unlike	*SS*	*MS*	*F*	*Significance F*	
Regresion	1	1.3180E+13	1.3180E+13	97.996477	**4.519547E-12**	
Residues	38	5.1108E+12	1.3449E+11			
Total	39	1.8291E+13				
	Coefficients	*Error mean*	*t Stat*	*p Value*	*Low 95%*	*Up 95%*
Limit	-1809144.346509	365670.350627	-4.947473	1.564156E-5	-2549405.263946	-1068883.429073
pay	176.690833	17.848790	9.899317	**4.519547E-12**	140.557846	212.823819

Table 3 MS Excel regression analysis

The value of coefficient of determination is sufficient enough and the model explains the dependence of data.As in previous cases, the regression model improved by gradually adding the variables of unemployment and exchange rates. Based on the values of Table 4 the variable from the model excluded employment, the parameter $\beta_1 = 0.916166$.

RESULT						
Regression statistics						
Correlation coefficient R	0.848920					
The coefficient determination R^2	**0.720666**					
Adjusting coefficient deternation R^2	0.705567					
Error mean	371601.486336					
Observation	40					
ANOVA						
	Unlike	*SS*	*MS*	*F*	*Significance F*	
Regresion	2	1.318155E+13	6.590774E+12	47.728912	**5.666253E-11**	
Residues	37	5.109244E+12	1.380877E+11			
Total	39	1.829079E+13				
	Coefficients	*Error mean*	*t Stat*	*p Value*	*Low 95%*	*Up 95%*
Limit	-1869798.193264	681758.469414	-2.742611	0.009339	-3251172.054919	-488424.331610
pay	177.662530	20.276733	8.761892	**1.480027E-10**	136.577968	218.747092
unymploed	114.641747	1081.670831	0.105986	**0.916166**	-2077.031520	2306.315015

Table 4 MS Excel regression analysis

For complete model of exchange rate CZK / EUR was a slight improvement in the coefficient of determination $R^2 = 0.764326$, and the parameter is statistically significant because the variable is included and the model will have the form (3) sp - value for β_1 *0.047701* and p - value for β_2 *0.012657*.

Regression statistics						
Correlation coefficient R	0.874258					
The coefficient determination R^2	**0.764326**					
Adjusting coefficient deternation R^2	0.751587					
Error mean	341327.226379					
Observation	40					
ANOVA						
	Unlike	*SS*	*MS*	*F*	*Significance F*	
Regresion	2	1.398013E+11	6.990067E+12	59.998371	**2.442053E-12**	
Residues	37	4.310658E+12	504275467.666E			
Total	39	1.829079E+13				
	Coefficients	*Error mean*	*t Stat*	*p Value*	*Low 95%*	*Up 95%*
Limit	3625648.897594	2101564.892065	1.725214	0.092834	-632526.014358	7883823.809546
pay	81.675678	39.880750	2.047998	**0.047707**	0.869604	162.481753
CZK/EURO	-124653.549786	47565.660425	-2.620663	**0.012657**	"-221030.731694	-28276.367878

Table 5 MS Excel regression analysis

Also refine the model using a variable exchange rate CZK / USD for the coefficient of determination was increased to 0.773751. There was, however, for including this variable in the regression model, the parameter is β_1

appears to be statistically insignificant (p-rank of the individual t-test 0.228691). Factor CZK / USD will be removed for reasons multicollinearity (CZK / EUR and CZK / UDS correlation coefficient 0.872949038).

Regressive model of dependence between factors will have the form (see Table 5):

$$HOTEL\ _{4*} = 3\,625\ \ 649 + 81.68 * pay - 12\ \ 465\ CZK\ / EUR \tag{3}$$

The evaluation of the relationship between the number of guests in hotels and other wages were created model (4, see Table 6). Given the very low leakage, $R^2 = 0.155951$ is needed to improve the model by adding another variable.

RESULT						
Regression statistics						
Correlation coefficient R	0,394906					
The coefficient determination R^2	**0.155951**					
Adjusting coefficient deternation R^2	0.133739					
Error mean	600495.599773					
Observation	40					
ANOVA						
	Unlike	*SS*	*MS*	*F*	*Significance F*	
Regresion	1	2.5317694E+12	2.531768E+12	7.021087	**0.011674**	
Residues	38	1.370261E+13	3.605950E+11			
Total	39	1.6234377E+13				
	Coefficients	*Error mean*	*t Stat*	*p Value*	*Low 95%*	*Up 95%*
Limit	4859966.609408	598752,247725	8,116824	**8.008878E-10**	3647856.06378	6072077.155036
pay	-77.440539	29.225786	-2.649733	**0,011674**	-136.605049	-18.276030

Table 6 MS Excel regression analysis

$$HOTEL\ _{next} = 4\ 859\ 967\ -77.4405 * pay \tag{4}$$

In determining the dependencies between variables unemployment and unemployment reached a value of regression model $R^2 = 0.218232$ ($adj\ R^2 = 0.175974$), where p-value of the individual t-test parameter $\beta_1 = 0.0943$ factor will be excluded from the model unemployment. Inclusion of the exchange rate CZK / EUR with values $R^2 = 0.230027$ (adj $R^2 = 0.188407$ and the value of the parameter β1 = 0.0676) and the exchange rate CZK / USD with the values of $R^2 = 0.196826$ ($adj\ R^2 = 0.153412$ and value of the parameter $\beta_1 = 0.1478255$) was elimination of the variables from the regression model and the dependence between the number of guests in hotels and other individual variables has the form (3)

Rating based on number of guests in boarding houses will be a factor in wage determination coefficient values of 0.46660 (adj. $R^2 = 0.4423$) with the β_1 *1.7619 E-06* included in the model, which will have the form:

$$HOTEL\ \ Boarding\ \ = 2\ 200\ 971.1\ -\ 62.12 * pay \tag{5}$$

Unemployment has been added to the model in order to improve its quality. This variable was also included in the model ($R^2 = 0.482938$), p - parameter *value 0.17807 β1*, appears as a parameter statistically insignificant. To further improve the model variables were gradually added in the case of exchange rate CZK / EUR was to create a model with the values of $R^2 = 0.5743$ *(R2 adj = 0.5513)* and p - values for the parameters $\beta_1 = 0.0000027$ and $\beta_2 = 0.002827$ has the form (Table 7):

$$HOTEL\ \ Boarding\ \ = 6\ \ 138\ \ 606.28\ -130.96 * pay - 90\ \ 314.42 * CZK\ / EUR \tag{6}$$

Adding the values of the variable CZK / USD has been improvement in the coefficient of determination, p - values for each parameter, they are statistically significant variables CZK / EUR and CZK / UDS, however, exhibit multicollinearity (correlation coefficient 0.872) and therefore the variable CZK / USD disabled regression model and linear model relationship will have the form (6)

RESULT						
Regression statistics						
Correlation coefficient R	0.757843					
The coefficient determination R^2	**0.574326**					
Adjusting coefficient deternation R^2	0.551317					
Error mean	202598.681776					
Observation	40					
ANOVA						
	Unlike	*SS*	*MS*	*F*	*Significance F*	
Regresion	2	2.049072E+12	1.024536E+12	24.960536	**1.373783E-07**	
Residues	37	1.518710E+12	41046225857.2908			
Total	39	3.567781E+12				
	Coefficients	*Error mean*	*t Stat*	*p Value*	*Low 95%*	*Up 95%*
Limit	6138606,28	1247407.894516	4.921090	1.799687E-05	3611117.825330	8666094.734638
pay	-130,9596302	23.671676	-5.532335	**2.699708E-06**	-178.923000	-82.996260
CZK/EURO	-90314,42071	28233.142144	-3.1988801	**0.002827**	-147520.200085	-33108.641330

Table 7 MS Excel regression analysis

This regression function was then examined in terms of heteroscedasticity using the Spearman correlation test sequence, using the Durbin-Watson test was examined first order autocorrelation. In both cases, the function appears to be satisfactory. The perimeter of the regression function is also well suited in multicollinearity.

Conclusion

As of model creation is evident in all models is lower or higher dependence on number of guests in accommodation establishments analyzed individually to pay. Probably the amount of income affects the choice of quality hotel accommodation. Already Lim (1997) confirmed in their study of tourism demand model, showed that economic factors such as income, relative prices, transportation costs and exchange rates are key factors in tourism demand [4].

In case of increase in wages only by CZK 1 occurs in 5 star hotels increased by 26 guests. In contrast, other hotels in unit wage increase causes a decrease of 77 non resident guests.

If the number of guests at hotels and guest houses 4 star depends not only on wages but also on the CZK / EUR. With the increase in wages and also to increase the exchange rate if there is a 4 star hotel to a decline in the number of guests in the case of pensions is similar.

The model shows that the movement of wages guests respond by changing preferences on the level of housing services and look for when a better salary as a stimulus for the use of a higher category of accommodation services. Increased exchange rate (devaluation of the Czech crown) can change the preferences of foreign guests, ie, increasing demand for higher categories of accommodation facilities.

References

[1] *Czech Statistical Office.* [On-line 2012] *Available from* http://www.czso.cz/csu/redakce.nsf/i/cru_cr , online 2012.

[2] Hušek, Roman, Pelikán, Jan.: *Aplikovaná ekonometrie: teorie a praxe.* Professional Publishing, Praha, 2003, 263.

[3] *Kurzy.cz Unemployment in the CR. Available from*: <http://www.kurzy.cz/makroekonomika/nezamestnanost/, online 2012

[4] Surugiu. C, Dincă ,A., Micu, D. *Tourism Destinations Vulnerable to Climate Changes: An Econometric Approach on Predeal Resort* //www.upg-bulletin-se.ro/archive/2010-1/13.%20Surugiu,%20Dinca,%20Micu.pdf , Romania, online 2010, 111-120.

[5] Wonnacott, T, H, Wonnacott R.,J.: *Statistika pro obchod a hospodářství.* Victoria Publishing, Praha, 1997, 891.

Fractional polynomials analysis of relation between insured accident and selected risk factors

Jiří Valecký[1]

Abstract. The paper is devoted to the examination of impact of selected risk factors on the insured accidence in motor hull insurance in the Czech Republic. In compliance with these risk factors, it is possible to determine the probability of insured accident for each policyholder and also to differentiate the paid insurance premiums which should corresponds to the undertaken risk. Therefore the impact of risk factors on the probability of insured accidence should be determined the most precisely. In this purpose, there are considered many joint variables, for instance volume and performance of engine, age of car, age of policyholder, and others. Firstly, using general univariable logistic regression, the most statistically significant risk factors are identified and it is determined how they affects occurrence of insured accident. It is also verified if the relation between them is linear. In the case of non-linearity identified, the most appropriate fractional polynomial function is obtained in order to improve fitting data. Finally, the non-linear functions are subjected to the stability analysis and the resulted linear and non-linear relations are interpreted.

Keywords: logit model, fractional polynomial, stability analysis, motor hull insurance, insurance premium, insured accident.

AMS Classification: 62 J 12, 62 P 05
JEL Classification: C31, C58, G22

1 Introduction

Insurance premium rate is determined per monetary unit in accordance with undertaken risk. In other words, the more risky client should pay higher premiums. This trend has been already observed for many years already. It is very common that the insurers set the premium in motor hull insurance in compliance with the volume of an engine or according to the size of district where the client lives. Moreover, some insurers respect even client´s age. To differentiate premiums in such way, the precise determining the relations between given particular risk factors and the outcome (insured accident) is crucial.

For the purpose of evaluating the undertaken risk represented by of insured accident probability, one can employ several models based on the GLM family model, see [1]. Nevertheless, the binomial models or count models are applied here, see [2] for instance. Some researchers utilize the advantages of both models and concentrate on the hurdle models representing mostly the combination of logit and some count (Poisson or negative-binomial) model designed firstly by [3].

It is necessary to note here that modelling relation between a risk factors and the outcome may suffer many imperfections resulted from excessive simplification. The simplification which we focus on is assumption of linear logit which may lead to the excessive distortion of relation between risk factor and the outcome. Due to this fact we may encourage to suppose that some risk factor affects the outcome differently for various values and thus the relation between them is necessary to model via non-linear function such as fractional polynomials (FPs), see [4] or [5].

The FPs are quite easy to use and interpret. In spite of it they suffer several imperfections. First and foremost, a selection of FPs is generally sensible to outliers and to data used to fitting the outcome. While the first problem represents the fact that one covariates affects the FPs, see [6], the latter incurs a lack of data fit when the estimated model is transferred to another data sample, see [7]. However, the high leverage does not necessary imply a large effect on fitting the outcome and on selecting FPs. Excluding the outlier observation or covariate pattern is needed to assess if the value may be considered as influential. Thus, the jackknife method, [8], is applied to FPs selection procedure and changes of FPs are examined. On the other hand, to assess the stability of estimated FPs is based on resampling and reestimation of the model (bootstrapping). All variously changing powers and degrees of FPs indicates the function instability. Moreover, allowing this type of uncertainty within FPs selection enlarge the confidence interval of the predicted outcome.

[1] VŠB-TU Ostrava, Faculty of Economics, Department of Finance, Sokolská tř. 33, 701 21, jiri.valecky@vsb.cz.

The aim of this paper is to assess the impact of selected risk factors on the insured accident probability and analyze the non-linear relation between them. The paper is organized as follows. The general logistic regression and its extension by incorporating fractional polynomials are described in Section 2. Section 3 is focused on the empirical examination of selected risk factors and Section 4 concludes the paper.

2 Identifying effect of risk factor on insured accident

Next, for the purpose of quantifying relation of risk factors to the insured accident, we focus on the logistic regression and fractional polynomials.

2.1 Linear logistic regression model

Consider a binary variable Y_i characterizing the occurrence of insured accident for given policyholder, thus

$$Y_i = \begin{cases} 1 & \text{if insured accident occurs,} \\ 0 & \text{otherwise, for } i = 1,...,N, \end{cases} \tag{1}$$

where N is the number of policyholders and each client is characterized by the vector of individual K risk factors $\mathbf{x}_i = (1, x_{1i}, x_{2i}, ..., x_{Ki})$.

The probability of an insured accident for a given policyholder, $P_i = P(Y_i = 1)$, is possible to express, on the basis of its characteristic vector $\mathbf{x}_i$, as a function $F(\boldsymbol{\beta}; \mathbf{x}_i)$ which is monotonically increasing $F'(\boldsymbol{\beta}; \mathbf{x}_i) \geq 0$ and has a domain of definition $(-\infty, +\infty)$ and a range $(0,1)$. Thus, it holds that $F(-\infty) = 0$ and $F(+\infty) = 1$, the probability function can be written in the form of

$$P_i = F(\boldsymbol{\beta}; \mathbf{x}_i), \tag{2}$$

where $\boldsymbol{\beta}$ is vector of parameters $(\beta_0, \beta_1, ..., \beta_K)$.

These properties are satisfied by the cumulative distribution function of the logistic distribution

$$P_i = P(Y_i = 1) = F(\boldsymbol{\beta}; \mathbf{x}_i) = \frac{e^{\boldsymbol{\beta}'\mathbf{x}_i}}{1 + e^{\boldsymbol{\beta}'\mathbf{x}_i}} = \frac{1}{1 + e^{-(\boldsymbol{\beta}'\mathbf{x}_i)}} \tag{3}$$

which is also a function of the probability that insured accident occurs. The probability that the accident does not occur can be written as

$$1 - P_i = P(Y_i = 0) = 1 - F(\boldsymbol{\beta}; \mathbf{x}_i) = \frac{1}{1 + e^{\boldsymbol{\beta}'\mathbf{x}_i}}. \tag{4}$$

The ratio of probabilities (3) and (4) is referred to as odds and it takes the form of

$$\frac{\pi}{1 - \pi} = \frac{P(Y_i = 1)}{P(Y_i = 0)} = e^{\boldsymbol{\beta}'\mathbf{x}_i} \tag{5}$$

and the logarithm of (5) is termed logit or log-odds, thus

$$\ln\left[\frac{\pi}{1 - \pi}\right] = \boldsymbol{\beta}'\mathbf{x}_i = g(\mathbf{x}_i). \tag{6}$$

2.2 Fractional polynomial function

Equation (6) may not necessary be linear. The non-linearity can be dealt with in that way that the predictor is converted into categorical factors but this procedure incorporates a problem arisen from making a cutpoints. The individuals close to but on opposite sides of the cutpoint are characterized very different rather than very similar, see [9]. Let's define the fractional polynomial function and rewrite the Equation (6) into the form of

$$g(x) = \beta_0 + \sum_{j=1}^{J} \beta_{1j} F_j(x_1) + \beta_2 x_2 + ... + \beta_K x_K, \tag{7}$$

where $F_j(x_1)$ is a particular type of power function. The power p_j could be any number, but [4] restricts the power to be among the set $S \in \{-2; -1; 0,5; 0; 0,5; 1; 2; 3\}$, where 0 denotes the log of the variable. The remaining functions are defined as

$$F_j(x_1) = \begin{cases} x_1^{p_j}, & p_j \neq p_{j-1} \\ F_{j-1}(x_1)\ln(x_1), & p_j = p_{j-1}. \end{cases} \tag{8}$$

The identification and comparison of the most appropriate FPs is made by closed test procedure, [10], which is generally preferred over the sequential procedure, [4].

3 Evaluating effects of selected risk factors

In this section, nonlinear impacts of given risk factors on the insured accident are examined. For this purpose, we used a data sample encompassing characteristics of policyholders in motor-hull insurance portfolio during the year 2008 (61 900 of insurance policies). In our study, we consider several continuous variables, i.e. age of a car (*agecar*) volume (*volume*) and performance of the engine (*kw*), age of a policyholder (*ageman*), number of citizens in a region (*nocit*) and average age in the region (*avgagereg*).

First and foremost, using the univariable logistic regression, the statistically significant predictors are identified and their linear effect on the outcome is verified by testing against the fractional polynomials considered at maximum second degree. Afterwards, if the non-linear effects are indicated, the FPs being statistically significant are estimated and the appropriateness of the FPs is assessed by residual analysis. Finally, because of the fact that the FPs are sensitive to the influential observations and to the sample used for estimation, we conduct a stability analysis to examine the stability of selected FPs.

3.1 Examination of non-linearity

Using the closed test procedure, the most appropriate FPs of the first and second degree is estimated and tested against each other and against the linear function. The results are recorded in the next table, where the second column informs about the statistical significance of variable inclusion in the model. The third and fourth columns report about the superiority of FP2 over the linear and FP1. Lastly, there are recorded estimated powers.

Variable	P-value for testing			Powers selected
	Inclusion	FP2 vs linear	FP2 vs FP1	
agecar	<0.001	0.097	0.043	1
volume	<0.001	<0.001	<0.001	0.5; 1
kw	<0.001	<0.001	<0.001	0.5; 1
ageman	<0.001	<0.001	<0.001	-2, -2
nocit	<0.001	0.617	0.510	1
avgagereg	<0.001	0.006	0.003	-2; -2

Table 1 Univariable analysis

It is obvious on the basis of p-values that all variables may be used for the purpose of insured accident modelling. Focusing on the decision about employing FPs, we can recommend FP2 for *volume*, *kw*, *ageman* and *avgagereg*. On the contrary, modelling *agecar* and *nocit* by FP2 is not statistically better and linear function (power selected = 1) is preferred at 5% level. The shapes of all functions are depicted in next figure to interpret the non-linear relation between the risk factor and the outcome.

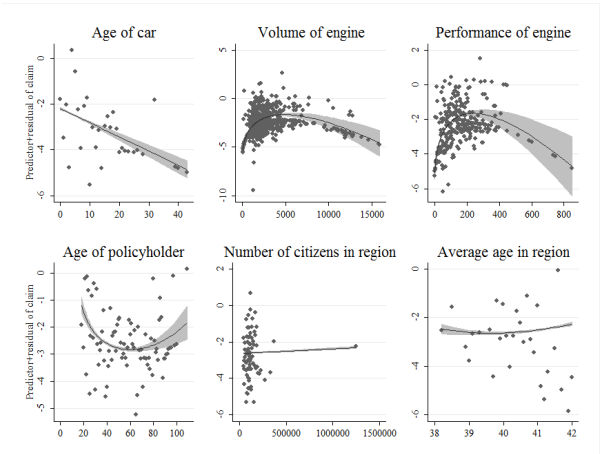


Figure 1 Behavior of functions: residuals and prediction with 95% CI

We can see that the *agecar* and *nocit* are modeled by a linear function, while the polynomials are used to model the other variables. On the basis of the figure above, we can conclude how the selected risk factors influ-

ence the outcome. When the higher the age of car is, the smaller probability of insured accident is. The impact of *volume* and *kw* on the outcome (logit or insured accident probability) is very similar, i.e. the insured accident probability gets higher as both variables increases and descends from some given value. In terms of the impact risk factor on the outcome, effect of *ageman* is also interesting. According to the shape of function, the lower the probability is when the *ageman* is increasing and then gets higher. The behavior of *avgagereg* is similar to the *ageman*. It results from the fact that the drivers around the 40 are the least risky drivers and therefore the region with lower average age indicates the least insured accident probability. Finally, we should note here that *nocit* should be probably modeled as categorical variable or offset rather than continuous variable and only *ageman* or *avgagereg* should be probably used in a multivariable model.

To assess the fact if the FPs improve the model fitting, we compare smoothed Pearson residuals of linear and FP functions for all variables which it is reasonable for, i.e. variables modeled by linear function and recommended as categorical are excluded, see next figure.

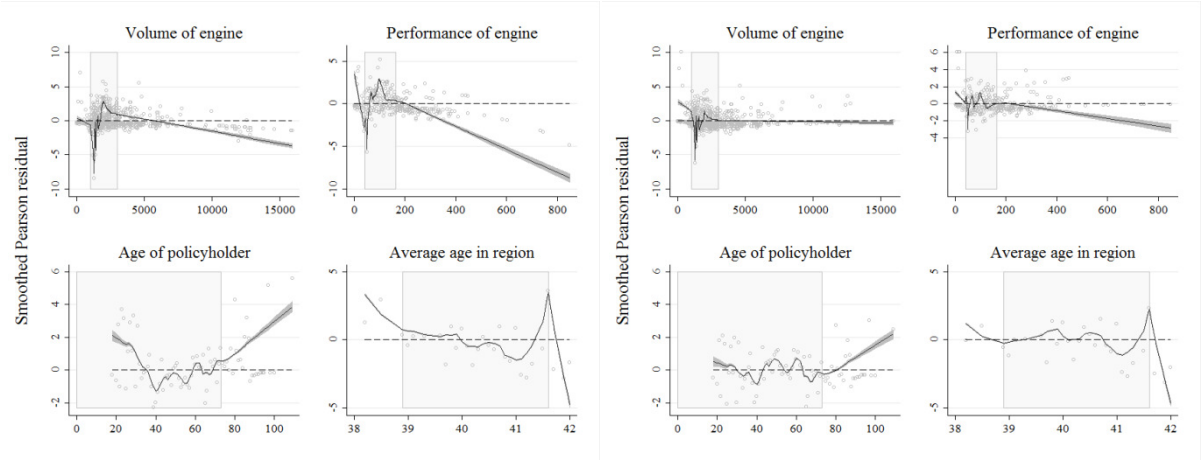


Figure 2 Smoothed Pearson residuals. Left: linear function. Right: FP2 functions

Well fitted observations are indicated by the smoothed residuals close to zero. From the figure above it is apparent that data are fitted better when the relations are modeled via FPs. The lack of fit was reduces in 95 % CI (gray rectangle) of all observations and even outside this region (for outliers).

3.2 Stability analysis of FPs

In the next step, we focus on identifying the influential covariates and in the second step we verify the stability of FPs selected. In the next figure, the calculated deviance difference between linear and FP2 are plotted against the excluded covariate pattern. If the difference is smaller than χ^2 threshold, selection of FP2 depends on this covariates. The plot is amended by the size of circle indicating the covariate frequency to reveal if the variable has several or many observations. The figure 3 gives the evidence that linear function for *ageman* is preferred over the FP2 because excluding all of covariate patterns (except one) results in linear function in FP selection procedure. Using FP2 to model the effect of *volume*, *kw* and *ageman* on the outcome is recommendable because none of the covariates affect the FP selection. Finally, selection of FP2 for modelling *avgagereg* is highly dependent on including the values 41.6 and a little on 38.5.

To assess the stability of selected FPs, we draw one hundred bootstrap sample of size 15 000 and apply the closed-test procedure to determine the most appropriate FPs. For each bootstrap sample is selected the most appropriate FP and the results are depicted in the next figure 4.

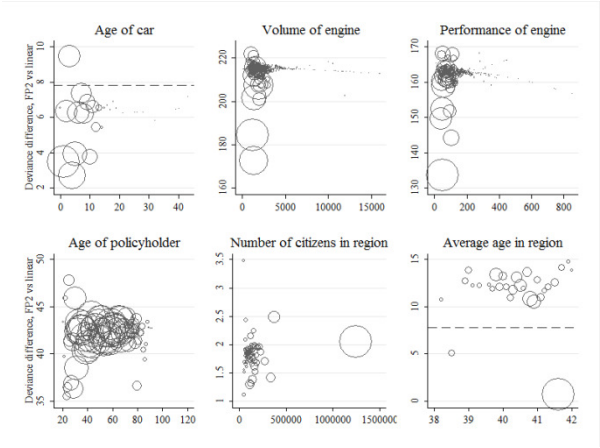


Figure 3 Influential points. Deviance difference between FP2 and linear function for each risk factor excluding each of covariate pattern in turn (circles are proportional to pattern frequencies). Horizontal line at 7.81 represents the χ^2 threshold for significance of FP2 versus linear at the 5% level.

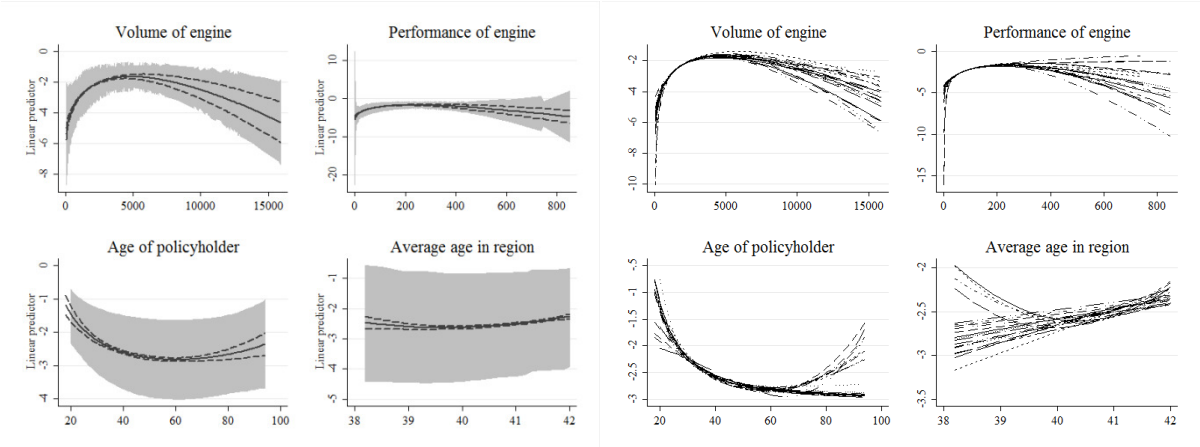


Figure 4 Bootstrap analysis (100 replications) of FPs: Left: Bootstrapped functions with CI: solid lines, means; dashed lines, 95% CI ignoring uncertainty due to FPs selection; gray area, bootstrapped 95% CI. Right: random set of 20 fitted FPs on bootstrap replications.

The predicted outcome can takes any value in 95% CI represented by the dashed line. However, this interval does not incorporate the uncertainty arisen from the selection of FPS using different sample. Therefore the bootstrapped CIs are wider (gray area). We can see that wide CI for *ageman* and *avgagereg* indicates also the high variability of degree and powers of FP. It is more evident in the figure on the right where the random set of 25 FPs out of the 100 bootstrap replications is depicted. We can mention that modelling of *avgagereg* by FP is unstable because of the linear function was selected within some bootstrap samples, probably depending on including the influential covariates, see above. The shape of FP for *ageman* also varies however the descent of probability with increasing age is still apparent. On the contrary, if the probability gets higher from given age of the driver is arguable. The shape of FP for *volume* and *kw* seems to be stable (except for the hook for small *kw*).

4 Conclusion

The paper was devoted to the analyzing impact of selected risk factors on insured accident in given motor hull insurance portfolio. Firstly, the significance of continuous predictors was evaluated within univariable logistic regression and the linear relation was verified. The revealed non-linearity was handled with the fractional polynomials of second degree and the relations were described and interpreted.

On the basis of conducted study, we found all considered continuous predictors to be statistically significant. The relation between age of car and the outcome may be approximated by linear function due to the many influential observations. Moreover, the number of citizens (also average age in region) should be converted into a categorical variable or should be used as an offset in the model rather than a continuous predictor. For others variables the fractional polynomials are recommended. In addition to the results of statistically significance at

the 95% confidence interval, the observations including outliers were fitted better and no other influential observations were identified.

We also revealed that the higher the age of man is, the smaller probability of insured accident is. However, the trend has changed for age above 60 approximately. As the stability analysis confirmed, the latter conclusion is discussable and should be subjected to further analysis because the fractional polynomial selected in the initial univariable analysis was unstable and monotonic decreasing FP was estimated on some bootstrap samples. The impact of volume and performance of the engine on the outcome was very similar, i.e. the insured accident probability gets higher as both variables increases and descends from some given value. The stability of these functions was confirmed except for the hook for low performance.

Acknowledgements

This paper was solved within the project P403/12/P692 Managing and modelling of insurance risks within Solvency II.

References

[1] Altman, D.G., and Andersen, P.K.: Bootstrap investigation of the stability of a Cox regression model. *Statistics in Medicine* **8** (1989), 771–783.
[2] Cameron, A.C., and Trivedi, P.K.: Econometric models based on count data: Comparisons and applications of some estimators. *Journal of Applied Econometrics* **1** (1986), 29–53.
[3] Marcus, R., Peritz, E., and Gabriel, K.R.: On closed test procedures with special reference to ordered analysis of variance. *Biometrika* **76** (1976), 655–660.
[4] Mosteller, F., and Tukey, J.W.: *Data analysis and regression*. Addison-Wesley, New York, 1977.
[5] Mullahy, J.: Specification and testing of some modified count data models. *Journal of Econometrics* **33** (1986), 341–365.
[6] Nelder, J.A., and Wedderburn, R.W.M.: Generalized linear models. *Journal of the Royal Statistical Society A* **135** (1972), 370–384.
[7] Royston, P., and Altman, D.G.: Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling (with discussion). *Applied Statistics* **43** (1994), 429–467.
[8] Royston, P., Ambler, G., and Sauerbrei, W.: The use of fractional polynomials to model continuous risk variables in epidemiology. *International Journal of Epidemiology* **28** (1999), 964–974.
[9] Royston, P., and Sauerbrei, W.: Stability of multivariable fractional plynomial models with slection of variables and transformations: a bootstrap investigation. *Statistics in Medicine* **22** (2003), 639–659.
[10] Royston, P., and Sauerbrei, W.: *Multivariable Model-building*. John Wiley & Sons, Chichester, 2008.

Determinants of commercial banks' liquidity in Poland

Pavla Vodová[1]

Abstract. The aim of this paper is to identify determinants of liquidity of Polish commercial banks. The data cover the period from 2001 to 2010. The results of panel data regression analysis showed that bank liquidity is strongly determined by overall economic conditions and dropped as a result of financial crisis, economic downturn and increase in unemployment. Bank liquidity decreases also with higher bank profitability, higher interest rate margin and bigger size of banks. On contrary, bank liquidity increases with higher capital adequacy, inflation, share of non-performing loans and interest rates on loans and interbank transaction.

Keywords: liquidity ratios, panel data regression analysis, Polish commercial banks.

JEL Classification: C23, G01, G21
AMS Classification: 62P20

1 Introduction

During global financial crisis, many banks struggled to maintain adequate liquidity. In order to sustain the financial system, unprecedented levels of liquidity support were required from central banks. Even with such extensive support, a number of banks failed, were forced into mergers or required resolution [4], [16]. The crisis showed the importance of adequate liquidity risk measurement and management.

It is evident that liquidity and liquidity risk is very up-to-date and important topic. The aim of this paper is therefore to identify determinants of liquidity of commercial banks in Poland.

The structure of the paper is following. After introduction as a first chapter, second chapter characterizes methods of bank liquidity measurement. Chapter 3 describes methodology and data used. Chapter 4 contains results of the analysis. Last chapter captures concluding remarks.

2 Methods of bank liquidity measurement

Liquidity is the ability of bank to fund increases in assets and meet obligations as they come due, without incurring unacceptable losses [3]. Liquidity risk arises from the fundamental role of banks in the maturity transformation of short-term deposits into long-term loans. It includes two types of risk:

- Funding liquidity risk is the risk that the bank will not be able to meet efficiently both expected and unexpected current and future cash flow and collateral needs without affecting either daily operations or the financial condition of the firm.
- Market liquidity risk is the risk that a bank cannot easily offset or eliminate a position at the market price because of inadequate market depth or market disruption.

There are three mechanisms that banks can use to insure against liquidity crises [2]:

- Banks hold buffer of liquid assets on the asset side of the balance sheet. A large enough buffer of assets such as cash, balances with central banks and other banks, debt securities issued by governments and similar securities or reverse repo trades reduce the probability that liquidity demands threaten the viability of the bank.
- Second strategy is connected with the liability side of the balance sheet. Banks can rely on the interbank market where they borrow from other banks in case of liquidity demand. However, this strategy is strongly linked with market liquidity risk.
- The last strategy concerns the liability side of the balance sheet, as well. The central bank typically acts as a Lender of Last Resort to provide emergency liquidity assistance to particular illiquid institutions and to provide aggregate liquidity in case of a system-wide shortage.

Liquidity risk can be measured by two main methods: liquidity gap and liquidity ratios. The liquidity gap is the difference between assets and liabilities at both present and future dates. At any date, a positive gap between assets and liabilities is equivalent to a deficit [5].

Liquidity ratios are various balance sheet ratios which should identify main liquidity trends. These ratios reflect the fact that bank should be sure that appropriate, low-cost funding is available in a short time. This might

[1] Silesian University in Opava, School of Business Administration in Karviná, Department of Finance, Univerzitní nám. 1934, 733 40 Karviná, vodova@opf.slu.cz.

involve holding a portfolio of assets than can be easily sold (cash reserves, minimum required reserves or government securities), holding significant volumes of stable liabilities (especially deposits from retail depositors) or maintaining credit lines with other financial institutions.

Various authors like [1], [2], [6], [8], [9], [11], [12], [13], [14] or [15] provide various liquidity ratios. For the purpose of this research we will use for evaluation of liquidity positions of Polish commercial banks following four different liquidity ratios (1) – (4):

$$L1 = \frac{liquid\ assets}{total\ assets} \cdot 100 \tag{1}$$

The liquidity ratio *L1* should give us information about the general liquidity shock absorption capacity of a bank. Cash, balances with central banks and other banks, debt securities issued by governments and similar securities or reverse repo trades belong to liquid assets. As a general rule, the higher the share of liquid assets in total assets, the higher the capacity to absorb liquidity shock, given that market liquidity is the same for all banks in the sample. Nevertheless, high value of this ratio may be also interpreted as inefficiency. Since liquid assets yield lower income liquidity bears high opportunity costs for the bank. Therefore it is necessary to optimize the relation between liquidity and profitability.

The liquidity ratio *L2* uses concept of liquid assets as well. However, this ratio is more focused on the bank's sensitivity to deposits (we included deposits of households and enterprises).

$$L2 = \frac{liquid\ assets}{deposits} \cdot 100 \tag{2}$$

The ratio *L2* measures the liquidity of a bank assuming that the bank cannot borrow from other banks in case of liquidity need. This is relatively strict measure of liquidity but it enables us to capture at least the part of the market liquidity risk. The bank is able to meet its obligations in terms of funding (the volume of liquid assets is high enough to cover volatile funding) if the value of this ratio is 100% or more. Lower value indicates a bank's increased sensitivity related to deposit withdrawals. The higher is the value, the higher is the capacity to absorb liquidity shock.

$$L3 = \frac{loans}{total\ assets} \cdot 100 \tag{3}$$

The ratio *L3* measures the share of loans in total assets. It indicates what percentage of the assets of the bank is tied up in illiquid loans. Therefore the higher this ratio the less liquid the bank is.

$$L4 = \frac{loans}{deposits} \cdot 100 \tag{4}$$

The last liquidity ratio *L4* relates illiquid assets with liquid liabilities. Its interpretation is the same as in case of ratio *L4*: the higher this ratio the less liquid the bank is. Lower values of this ratio means that loans provide by the bank are financed by deposits.

These liquidity ratios are still in common. It is possible to calculate them only on the basis of publicly available data from banks´ balance sheets and it is easy to interpret their values. Their disadvantage is the fact that they do not always capture all, or any of liquidity risk.

3 Methodology and data

As in case of our previous studies about determinants of liquidity of Czech and Slovak commercial banks [18] and [19], in order to identify determinants of liquidity of Polish commercial banks we use the panel data regression analysis. For each liquidity ratio, we estimate equation (5):

$$L_{it} = \alpha + \beta \cdot X_{it} + \delta_i + \varepsilon_{it} \tag{5}$$

where L_{it} is one of four liquidity ratios (*L1* – *L4*) for bank *i* in time *t*, X_{it} is vector of explanatory variables for bank *i* in time *t*, α is constant, β' is coefficient which represents the slope of variables, δ_i represents fixed effects in bank *i*, and ε_i means the error term.

It is evident that the most important task is to choose the appropriate explanatory variables. Although liquidity problems of some banks during global financial crisis re-emphasized the fact that liquidity is very important for functioning of financial markets and the banking sector, an important gap still exists in the empirical literature about liquidity and its measuring. Only few studies aim to identify determinants of liquidity, such as [6], [7], [10], [12] or [17]. The review of these studies can be found in [18] and [19]. These studies suggest that commercial banks' liquidity is determined both by bank specific factors (such as size of the bank, profitability, capital adequacy and factors describing risk position of the bank) as well as macroeconomic factors (such as different types of interest rates, interest margin or indicators of economic environment). It can be useful to take into account some other influences, such as the realization of financial crisis, changes in regulation or political incidents.

The selection of variables was based on above cited relevant studies. We considered whether the use of the particular variable makes economical sense in Polish conditions. For this reason, we excluded from the analysis variables such as political incidents, impact of economic reforms or the exchange rate regime. We also considered which other factors could influence the liquidity of banks in Poland. The limiting factor then was the availability of some data. Table 1 shows a list of variables which we have used in regression analysis.

variable	definition	source	estim. effect
CAP	the share of equity on total assets of the bank	annual rep.	+
NPL	the share of non-performing loans on total volume of loans	annual rep.	-
ROE	return on equity: the share of net profit on banks´ equity	annual rep.	-
TOA	logarithm of total assets of the bank	annual rep.	+/-
FIC	dummy variable for realization of financial crisis (1 in 2008 and 2009, 0 in rest of the period)	own	-
GDP	growth rate of gross domestic product growth (96499BPXZF... GDP volume % change)	IMF	+/-
INF	inflation rate: (96464..XZF...CPI % change)	IMF	+
IRB	interest rate on interbank transactions: (96460B..ZF... money market rate)	IMF	+
IRL	interest rate on loans: (96460P..ZF... lending rate)	IMF	-
IRM	difference between interest rate on loans (96460P..ZF... lending rate) and int. rate on deposits (96460L..ZF...deposit rate)	IMF	-
MIR	monetary policy interest rate repo rate: (96460...ZF... repo rate)	IMF	-
UNE	unemployment rate: (96467R..ZF...unemployment rate)	IMF	-

Table 1 Variables definition

We consider four bank specific factors and eight macroeconomic factors. As it can be seen from Table 1, we expect that three factors could have positive impact on bank liquidity, the rest of factors are expected to have negative impact on bank liquidity. Macroeconomic data were provided by International Financial Statistics of International Monetary Fund (IMF). Bank specific data were obtained from annual reports of Polish banks. We used unconsolidated balance sheet and profit and loss data over the period from 2001 to 2010. The panel is unbalanced as some of the banks do not report over the whole period of time.

	01	**02**	**03**	**04**	**05**	**06**	**07**	**08**	**09**	**10**
total number of banks	69	59	58	5	54	51	50	52	49	49
number of observed banks	26	29	33	35	36	33	32	32	30	27
share of observed banks on total assets (%)	71	74	89	85	85	83	81	80	78	75

Table 2 Data availability

Table 2 shows more details about the sample. As it includes substantial part of the Polish banking sector, we used fixed effects regression.

4 Results

We use an econometric package EViews 7. After tests of stationarity, we proceed with regression estimation. We estimate (6) separately for each of four defined liquidity ratios. We gradually change the content of the vector of

explanatory variables X_{it}. The aim is to find a model which has a high adjusted coefficient of determination and simultaneously the variables used are statistically significant. As it can be seen from following tables, results of the analysis suggest that liquidity ratios are determined by different factors.

If we measure liquidity with ratio *L1*, we find determinants of liquidity in Table 3. The explanatory power of this model is quite high and signs of coefficients correspond with our expectations. The positive influence of the share of capital on total assets is consistent with the assumption that bank with sufficient capital adequacy should be liquid, too. Liquidity is decreasing with the size of the bank. It seems that big banks insure against liquidity crises mainly by passive strategies: they rely on the interbank market or on a liquidity assistance of the Lender of Last Resort. This finding fully corresponds to the well known "too big to fail" hypothesis. If big banks are seeing themselves as "too big to fail", their motivation to hold liquid assets is limited.

Results also show negative impact of interest margin on bank liquidity. This is logical: increase in interest margin stimulates bank to focus more on lending activity and as a result, the share of liquid assets is decreasing. Inflation increases banks´ vulnerability to nominal values of loans provided to customers so during periods of inflation banks hold more liquid assets. The positive coefficient on GDP growth rate signals that according to our expectations, liquidity tends to be inversely related to the business cycle. Most borrowers want to take a loan during expansion when they have valuable investments projects. Banks which want to satisfy the growing demand for loans would face lower liquidity. During economic downturn, lending opportunities are not so good so banks hold higher share of liquid assets. As we have expected, financial crisis and bank liquidity is inversely related. Financial crisis could be caused by poor bank liquidity. However, the effect may be the opposite: financial crisis lead to poor bank liquidity. Financial crisis affects banks in two different ways. First, the volatility of important macroeconomic variables influences unfavorably the business environment of banks. Second, the instability deteriorates the business environment of borrowers; it can worsen their ability to repay the loans which can lead to a decline in bank liquidity. This is fully confirmed also by the last explanatory variable – increase in the rate of unemployment (which may be a result of financial crisis) has negative impact on bank liquidity.

L1			**L2**		
variable	coefficient	st. deviation	variable	coefficient	std. deviation
constant	168.890*	32.4178	constant	-19650.0*	2492.26
CAP	0.33120*	0.11462	CAP	1292.80*	82.8136
TOA	-7.45713*	2.06426	GDP(-1)	591.271**	247.149
IRM	-6.21024*	1.74657	ROE	-60.9919***	35.4199
INF	3.00102*	1.12920	IRL	475.884***	249.782
GDP	-3.09675*	0.99721			
FIC	-14.7652*	4.10656			
UNE	-2.68292*	0.70235			
adjusted R^2	0.678602		adjusted R^2	0.729348	
total observ.	290		total observ.	245	

Table 3 Determinants of liquidity measured by *L1* and *L2*. The starred coefficient estimates are significant at the 1% (*), 5% (**) or 10 % (***) level.

Table 3 shows also determinants of liquidity measured by the ratio *L2*. Explanatory power of the model is slightly higher. The share of liquid assets on bank deposits is determined by bank capital adequacy, profitability, interest rate on loans and growth rate of GDP in previous year. We have discussed the impact of capital adequacy above. The link between bank profitability and liquidity is consistent with standard finance theory which emphasizes the negative correlation of liquidity and profitability. Signs of other two coefficients do not correspond with our expectations. Although most studies assumed the negative link between business cycle and bank liquidity, the results show that the approach of Moore [12] can be applied on Polish banking sector. Positive sign of the coefficient signals that cyclical downturn should lower banks' expected transactions demand for money and therefore lead to decreased liquidity. Moreover, during expansionary phases, companies (which have higher profits) and households (which have higher income) might prefer to rely more on internal sources of finance and reduce the relative proportion of external financing and might reduce their debt levels. In recessions, households and corporations may increase their demand for bank credit in order to smooth out the impact of lower income and profits. Growth rate of gross domestic product is statistically significant with one year lag which is consistent with the fact that companies and households need some time for accumulating profits and savings. The

results show the positive link between interest rate on loans and bank liquidity, which correspond neither to our expectations nor to a standard economic theory. Banks probably focus more on the interest margin or it can highlight the fact that higher lending rates do not encourage banks to lend more. This is consistent with the problem of credit crunch and credit rationing.

Determinants of liquidity measured by the ratio *L3* are presented in Table 4. As high value of this ratio means low liquidity, these results have to be interpreted in reverse: positive sign of the coefficient means negative impact on liquidity and conversely. Explanatory power of the model is very high. Only three variables are statistically significant. The link between bank liquidity and interest rate margin and inflation were described above. Although we estimated negative influence of non-performing loans, results show the opposite effect. This could be a sign of prudent policy of banks: they offset the higher credit risk in previous years with lower lending activity and cautious liquidity risk management.

L3			**L4**		
variable	coefficient	st. deviation	variable	coefficient	std. deviation
constant	53.3822*	3.30393	constant	-22376.7	13811.0
NPL(-2)	-0.32658*	0.09088	IRB	2956.77*	1057.60
IRM	3.32758*	0.69906	GDP	4991.89**	2137.90
INF	-1.65585**	0.70152			
adjusted R^2	0.953329		adjusted R^2	0.621131	
total observ.	220		total observ.	290	

Table 4 Determinants of liquidity measured by *L3* and *L4*. The starred coefficient estimates are significant at the 1% (*), 5% (**) or 10 % (***) level.

Table 4 shows also determinants of liquidity measured by the last liquidity ratio *L4*. The explanatory power of this last model is slightly lower. The share of loans on bank deposits is determined by growth rate of GDP and by the level of interest rate on interbank transactions. As in case of liquidity ratio *L1* (and in opposite to liquidity ratio *L2*), the positive sign of the coefficient (and thus negative influence on bank liquidity) signals that liquidity is inversely related to the business cycle. The impact of interbank interest rate is not in accordance with our expectations but it is the same as in case of impact of interest rate on loans. It seems that overall lending activity (both with other banks and nonfinancial sector) and thus indirectly the liquidity of banks is not determined by the level of interest rates.

5 Conclusion

The aim of this paper was to identify determinants of liquidity of commercial banks in Poland. We have used the panel data regression analysis for four liquidity ratios. We consider four bank specific factors and eight macroeconomic factors and with the only exception of repo interest rate, all variables were at least in some models statistically significant. The results of models enable us to make following conclusions.

Bank liquidity is strongly determined by overall economic conditions and dropped as a result of financial crisis, economic downturn and increase in unemployment. Increase in interest rate margin and bank profitability lowers bank liquidity, too. Liquidity also decreases with the size of the bank: big banks rely on the interbank market or on a liquidity assistance of the Lender of Last Resort, small and medium sized banks hold buffer of liquid assets.

Increases in capital adequacy, inflation, share of non-performing loans and level of interest rate both on loans and interbank transaction have positive impact on bank liquidity.

Acknowledgements

This paper was prepared with financial support of Czech Science Foundation (Project GAČR P403/11/P243: Liquidity risk of commercial banks in the Visegrad countries).

References

[1] Andries, A.M.: A comparative analysis of performance and soundness indicators of the main Romanian banks. *Scientific Annals of the "Alexandru Ioan Cuza" University of Iasi: Economic Science Series* **LVI** (2009), 45–70.

[2] Aspachs, O., Nier, E., and Tiesset, M.: Liquidity, Banking Regulation and the Macroeconomy. Evidence on bank liquidity holdings from a panel of UK-resident banks. *Bank of England Working Paper* (2005).
[3] Basle Committee on Banking Supervision: *Principles for Sound Liquidity Risk Management and Supervision*. Bank for International Settlements, Basel, 2008.
[4] Basle Committee on Banking Supervision: *International framework for liquidity risk measurement, standards and monitoring*. Bank for International Settlements, Basel, 2009.
[5] Bessis, J.: *Risk Management in Banking*. John Wiley & Sons Inc., Chichester, 2009.
[6] Bunda, I., and Desquilbet, J.B.: The Bank Liquidity Smile Across Exchange Rate Regimes. *International Economic Journal* **3** (2008), 361–386.
[7] Fielding, D., and Shortland, A.: Political Violence and Excess Liquidity in Egypt. *Journal of Development Studies* **4** (2005), 542–557.
[8] Ghosh, S.: Credit Growth, Bank Soundness and Financial Fragility: Evidence from Indian Banking Sector. *South Asia Economic Journal* **11** (2010), 69–98.
[9] Jiménez, G., Ongena, S., Peydró, J., and Saurina, J.: Credit Supply: Identifying Balance-Sheet Channel with Loan Applications and Granted Loans. *ECB Working Paper Series* **1179** (2010).
[10] Lucchetta, M.: What Do Data Say About Monetary Policy, Bank Liquidity and Bank Risk Taking? *Economic Notes by Banca Monte dei Paschi di Siena SpA* **2** (2007), 189–203.
[11] Maechler, A.M., Mitra, S., and Worrell, D.: Decomposing Financial Risks and Vulnerabilities in Eastern Europe. *IMF Working Paper* **248** (2007).
[12] Moore, W.: How do financial crises affect commercial bank liquidity? Evidence from Latin America and the Caribbean. *MPRA Paper* **21473** (2010).
[13] Praet, P., and Herzberg, V.: Market liquidity and banking liquidity: linkages, vulnerabilities and the role of disclosure. *Banque de France Financial stability Review,* 2008, 95–109.
[14] Rychtárik, Š.: Liquidity Scenario Analysis in the Luxembourg Banking Sector. *BCDL Working Paper* **41** (2009).
[15] Tamirisa, N.T., and Igan, D.O.: Are Weak Banks Leading Credit Booms? Evidence from Emerging Europe. *Comparative Economic Studies* **4** (2008), 599–619.
[16] Teplý, P.: The Future Regulatory Challenges of Liquidity Risk Management. *World Academy of Science, Engineering and Technology* 73 (2011), 945–949.
[17] Valla, N., and Saes-Escorbiac, B.: Bank liquidity and financial stability. *Banque de France Financial Stability Review*, 2006, 89–104.
[18] Vodová, P.: Liquidity of Czech Commercial Banks and its Determinants. *International Journal of Mathematical Models and Methods in Applied Sciences* **6** (2011), 1060–1067.
[19] Vodová, P.: Liquidity of Slovak Commercial Banks and its Determinants. In: *Proceedings of the 13th International Conference on Finance and Banking* (Stavárek, D., and Vodová, P., eds.). Silesian University, Karviná, 2012, 487–494.

On problem of optimization under incomplete information

Petr Volf[1]

Abstract. In the paper we study consequences of incomplete information to uncertainty of results of stochastic optimization. Stochastic characteristics of optimized system are evaluated from observed data, moreover, the data may be incomplete. Namely, we consider the random censoring of observations frequently encountered in time-to-event (of lifetime) studies. The analysis of uncertainty will be based both on theoretical properties of estimated stochastic characteristics and on simulated examples.

Keywords: optimization, censored data, Fisher information, product-limit estimate.

JEL classification: C41, J64
AMS classification: 62N02, 62P25

1 Introduction

In problems of optimization under uncertainty we often rely on a probabilistic model of optimized system. Then the optimization task can, schematically, have the form

$$\varphi(F) = \inf_{\boldsymbol{y}} E_F \boldsymbol{C}(\boldsymbol{y}, \boldsymbol{v}), \tag{1}$$

where $\boldsymbol{C}$ is a cost function, $\boldsymbol{y}$ are input variables (observed), E_F stands for the expectation under distribution function F, and, finally, $\boldsymbol{v}$ is a random variable (or vector) with distribution function F. If F is known, we deal with a "deterministic" optimization. However, our information on probability distributions governing the system could be non-complete. Either, known distribution type depends on unknown parameters. Then, as a rule, the estimates of parameters are plugged into optimal solutions. Or, we have to employ nonparametric estimates, as is the empirical distribution function. Hence, our information on F is random and we have to analyze both possible bias and variability of obtained solution (compared to an ideal solution when F is known). Alternatively, we then can be interested in a kind of multi-objective optimization, minimizing simultaneously also variability (measured by variance, or certain inter-quantile range). Nevertheless, standard approach considers a solution of (1) and uses estimated characteristics instead of 'true' ones. An investigation of usage of empirical (estimated) characteristics in stochastic optimization problems started already in 70-ties. A set of papers is available, let us mention here just two: Dupačová and Wets (1984), and from more recent time Kaňková (2009) with brief overview and a number of other references.

The situation is even more complicated if the data available for estimation are not complete. We shall consider one special type of incompleteness, the random censoring from the right side. It is quite frequent in the analysis of demographic, survival or insurance data. The lack of information leads to higher variability (and, sometimes, to a bias) of estimates and, consequently, to higher uncertainty of optimal solutions.

The approaches to statistical data analysis in cases when the data are censored or even truncated are provided by a number of authors. Let us mention here works of C. Huber (e. g. Huber, 2000), with classification of deigns of censored and truncated data and with many references to papers dealing with specific methods of such data processing. The most of results were derived in the framework of statistical survival analysis and collected also in several monographs (cf. Kalbfleisch and Prentice, 2002).

The main objective of the present paper is to study the increase of uncertainty of results of optimization problem when the censoring is causing growing variability of estimates. We shall deal with both

[1]Institute of Information Theory and Automation, Prague 8, e-mail: volf@utia.cas.cz

parametric and non-parametric cases. To this end, certain theoretical properties of estimates under random right censoring will be recalled. In the next two parts, we shall consider the product-limit estimate as a generalization of the empirical distribution function, and then the maximum likelihood estimates of parameters when random right censoring is present. We shall compare properties of estimates with and without censoring, in nonparametric case (in Part 2) as well as in the case of estimated parameters (Part 3). Finally, we shall study the consequence of incomplete data to a certain problem of optimization, in Part 4. An example will deal with optimal maintenance schedule, properties of obtained 'quasi-optimal' solution will be illustrated with the aid of simulations.

2 Non-parametric case and product-limit estimate

Let us first recall the scheme of randomly right-censored data. Consider a random variable Y characterizing a random time to certain event, another random variable Z is a censoring variable, both are positive, continuous and mutually independent. Further, denote $f(y)$, $g(z)$, $F(y)$, $G(z)$, $\overline{F}(y) = 1-F(y)$, $\overline{G}(z) = 1 - G(z)$ density, distribution and survival functions of both variables. It is assumed that we observe just $X = \min(Y, Z)$ and $\delta = 1[Y \leq Z]$, i. e. δ indicates whether Y is observed or censored from right side. The data are then given as random sample $(X_i,\, \delta_i,\, i = 1, \ldots, N)$. Notice that the case without censoring is obtained when $G(z) \equiv 0$ on region where $F(y) < 1$.

A generalization of empirical distribution function is well known Kaplan–Meier "Product Limit Estimate" of survival function. Let us first sort (re-index) the data in increasing order, $X_1 \leq X_2 \leq \cdots \leq X_N$, then this estimator has the form

$$\overline{F}_N(t) = \prod_{i=1}^{N} \left(\frac{N-i}{N-i+1} \right)^{\delta_i \cdot 1[X_i \leq t]}. \tag{2}$$

Again, notice that when all $\delta_i = 1$, we obtain the empirical survival function. The following proposition is due Breslow and Crowley (1974).

Proposition: Let $T > 0$ be such that still $\overline{F}(T) \cdot \overline{G}(T) > 0$. Then random process

$$V_N(t) = \sqrt{N} \left(\frac{\overline{F}_N(t)}{\overline{F}(t)} - 1 \right) = \sqrt{N}\, \frac{F_N(t) - F(t)}{\overline{F}(t)} \tag{3}$$

converges, on $[0, T]$, when $N \to \infty$, to Gauss martingale with zero mean and variance function

$$C(t) = \int_0^t \frac{\mathrm{d}F(s)}{\overline{F}(s)^2\, \overline{G}(s)}. \tag{4}$$

It means that, in other words, $V_N(t)$ converges in distribution on $[0, T]$ to Brown process $\beta(C(t))$, or the process $D_N(t) = V_N(t)/(1+C(t))$ to Brownian bridge on $[0,\, C(T)/(1+C(T))]$. The asymptotic variance function can be estimated by its empirical version:

$$C_N(t) = \sum_{i=1}^{N} \frac{N\delta_i}{(N-i+1)^2} \cdot 1[X_i \leq t],$$

which is consistent uniformly on $[0, T]$.

For the case without censoring we obtain that $C(t) = F(t)/\overline{F}(t)$ and $D_N(t) = \sqrt{N}(F_N(t) - F(t))$ leading to standard Kolmogorov–Smirnov statistics. From (4) it is also seen that variance in the case with censoring (when $\overline{G}(t) \leq 1$) is larger than without it (i.e. when $\overline{G}(t) = 1$ on whole $[0, T]$). Further, it has been proved (see, for instance Robbins and Siegmund, 1970) that for $c,\, d > 0$ and sufficiently large T,

$$P\left(\sup_{0<t<T} |\beta(t)| < c + d \cdot t \right) \doteq 1 - 2\exp(-2cd).$$

Hence, if we take $c = d$ and "time" $C(t)$ instead t, we obtain that approximately

$$P\left(\sup_t |D_N(t)| > c \right) \doteq 2\exp(-2c^2).$$

In order to construct $1-\alpha$ band for $D_N(t)$ on $(0, C(T))$, we set $\alpha = 2\exp(-2c^2)$ and obtain critical value $c_\alpha = \sqrt{-\ln\frac{\alpha}{2}/2}$. In the case without censoring c_α is the distribution-free critical value for the Kolmogorov–Smirnov test, namely

$$P\left(\sup_t |F_N(t) - F(t)| \geq c_\alpha/\sqrt{N}\right) \doteq \alpha.$$

In the case with censoring, we have

$$D_N = \sqrt{N}\left(F_N(t) - F(t)\right)/\overline{F}(t)/(1 + C(t)),$$

hence, corresponding $1-\alpha$ confidence band for $F(t)$ depends on both F and G and its width is increasing for larger t. Namely, $1-\alpha$ borders for $|F_N(t) - F(t)|$ on $[0, T]$ are given as $c_\alpha/\sqrt{N} \cdot \overline{F}(t) \cdot (1 + C(t))$.

Example: Let us here, as an example, consider so called Koziol–Green model assuming that $\overline{G}(t) = \overline{F}(t)^a$, for some $a > 0$. Then

$$C(t) = \int_0^t \frac{\mathrm{d}F(s)}{\overline{F}(t)^{2+a}} = \frac{1}{\overline{F}(t)^{1+a}} - 1$$

and $\overline{F}(t) \cdot (1 + C(t)) = 1/\overline{F}(t)^a$. It tends to infinity with increasing t because $\overline{F}(t) \to 0$. A more concrete example is presented in Part 4.

3 Parametric estimates under censoring

In the present part we shall study the influence of censoring to precision of estimated parameters. It means that we assume that the type of distribution $F(y, \theta)$ of random variable Y is known and parameter θ is estimated by the method of maximum likelihood. The precision of estimation will be based on the Fisher information. It is defined as

$$I(\theta) = E\left(\frac{\mathrm{d}\ln L(\theta, X)}{\mathrm{d}\theta}\right)^2,$$

where $L(\theta, X)$ is the likelihood of θ based on random variable X. Naturally, if θ is multi-dimensional, we consider a vector of partial derivatives, What is important from our point of view, $I^{-1}(\theta)$ is also the asymptotic variance of $\sqrt{N}(\hat{\theta}_N - \theta)$, where $\hat{\theta}_N$ is the maximum likelihood estimate from random sample of extent N.

In the case of random right censoring, the log of likelihood (its part depending on θ), is $\ln L(\theta, X) = \delta \cdot \ln f(X) + (1-\delta) \cdot \ln \bar{F}(X)$. Hence, with notation $f' = \mathrm{d}f/\mathrm{d}\theta$ and $\bar{F}' = \mathrm{d}\bar{F}/\mathrm{d}\theta$,

$$E\left(\frac{\mathrm{d}\ln L(\theta, X)}{\mathrm{d}\theta}\right)^2 = \int_0^\infty \bar{G}(x)\left(\frac{f'(x)}{f(x)}\right)^2 f(x)\mathrm{d}x + \int_0^\infty \bar{F}(x)\left(\frac{\bar{F}'(x)}{\bar{F}(x)}\right)^2 g(x)\mathrm{d}x.$$

When the second integral is transformed with the aid of *per-partes*, we obtain that

$$I(\theta) = \int_0^\infty \bar{G}(x)\frac{(f'(x)\bar{F}(x) - f(x)\bar{F}'(x))^2}{f(x)\bar{F}((x)^2}\mathrm{d}x,$$

which is positive and is larger when $\bar{G}(x) \equiv 1$, i.e. when there is no censoring. Again, a more concrete comparison is presented within the example in Part 4.

4 Example of optimization problem

Let us consider the following rather simple example of optimization problem: A component of a machine has its time to failure Y given (modeled) by a continuous-type probability distribution with distribution function, density, survival function F, f, $\bar{F} = 1 - F$, respectively. The cost of repair after failure is C_1, the cost of preventive repair is $C_2 < C_1$. For the simplicity we assume that only complete repairs, 'renewals', are provided, i.e. after each repair the component is new (exchanged) or as new. Let τ be the (fixed) time from renewal to preventive repair.

Then, the mean time between renewals is $ET(\tau) = \tau \cdot P(Y > \tau) + \int_0^\tau t f(t)\mathrm{d}t = \int_0^\tau \bar{F}(t)\mathrm{d}t$, while the mean cost of one renewal equals $EC(\tau) = C_1 \cdot F(\tau) + C_2 \cdot \bar{F}(\tau)$. Our task is to find optimal τ minimizing

$$\min_\tau \phi_F(\tau) = \min_\tau EC(\tau)/ET(\tau). \tag{5}$$

In the sequel the lifetime distribution will be specified and we shall compare the deterministic solution provided F is known, and the variability of 'quasi-solutions' in cases when lifetime distribution is estimated, in parametric or non-parametric setting, from censored and non-censored data. Namely, let the distribution of Y be Weibull, with parameters $a = 100, b = 2$, i.e. its survival function is $\overline{F}(t) = \exp\left(-\left(\frac{t}{a}\right)^b\right)$, numerical characteristics are $EY \sim 89$, $std(Y) \sim 46$. Costs of repairs were fixed as $C_1 = 10$, $C_2 = 1$. When the distribution function F is known, there exists an unique optimal solution with $\tau^* = 33.64$ and minimal costs per time unit $\phi(\tau^*) = 0.061$.

In the next parts we provide a numerical study where it is assumed that the distribution of variable Y is estimated from data. In all from four cases (parametric or nonparametric case, without or with censoring) 100 samples of 100 observation Y_i are generated from the Weibull distribution mentioned above. In cases with censoring, they are censored by the censoring variables Z_i having uniform distribution on $[0, 250]$, hence with survival function $\overline{G}(z) = (250 - z)/250$ (value 250 corresponds roughly to 0.998 quantile of distribution of Y). The rate of censoring is then about 36 % $\sim EY/250$.

4.1 Parametric case

In the first part of the study the Weibull-type distribution was taken for granted, its parameters were estimated from generated samples of data, by the maximum likelihood method. Hence, 100 couples of estimates $a_m, b_m, m = 1, 2, ..., 100$ were obtained. They are displayed in Figure 1, left plot shows estimates from non-censored cases, the right plot corresponds to censored cases. From those values their means and sample standard deviations were computed. Simultaneously, we computed theoretical Fisher information I for both parameters and approximate standard deviations of estimates I-std$= \sqrt{1/I/N}$, for extent of sampled data $N = 100$. All these characteristics are collected in Table 1. Further, to each estimated couple of parameters, an optimal solution of (5), τ_m and $\phi(\tau_m)$, $m = 1, 2, ..., 100$, was computed. They are shown in Figure 2, again for non-censored (left) and censored cases (right). Table 1 contains also sample means and standard deviations of those 'optimal' τ_m and $\phi(\tau_m)$.

	non-censored: I	I-std	sample mean	sample std	censored: I	I-std	sample mean	sample std
a	$3.80 \cdot 10^{-4}$	5.13	100.3849	5.5523	$2.64 \cdot 10^{-4}$	6.15	100.0807	6.9541
b	0,409	0,154	2.0147	0.1585	0.287	0.187	2.0074	0.1680
τ^*			33.9970	1.9425			33.9070	2.3610
$\phi(\tau^*)$			0.0606	0.0071			0.0611	0.0070

Table 1: Comparison of theoretical and sample-based characteristics of estimates and optimal solutions

4.2 Nonparametric estimate of distribution function

Let us now imagine that we do not know the type of distribution of Y and therefore we estimate it with the aid of the product-limit estimator (i.e. as the empirical distribution function when censoring is absent). Figure 3 displays cloud of 100 estimates obtained from 100 generated samples, the cases without censoring are plotted in the left subplot, the right subplot shows estimates obtained from censored data. It is well seen how the variability in the right subplot increases for large times. Theoretically, if we take $\alpha = 0.05$ and $N = 100$, the half-width of 95% confidence band for 'true' distribution function, in the non-censored case, is given approximately as $c_\alpha/\sqrt{N} = 0.136$. As regards the censored data, function $C(t)$ (here defined on $[0, 250)$) has no analytical form, nevertheless, we can compute it numerically. We have seen in Part 3 that, at a given t, the half-width of $1-\alpha$ band is given as $d_{N,\alpha}(t) = c_\alpha/\sqrt{N} \cdot \overline{F}(t) \cdot (1 + C(t))$. We computed it at three points corresponding roughly to three quartiles of utilized Weibull distribution. Namely, at points $t = 55$, 85, 120 we obtained $d_{N,\alpha}(t) = 0.142$, 0.158, 0.195, respectively.

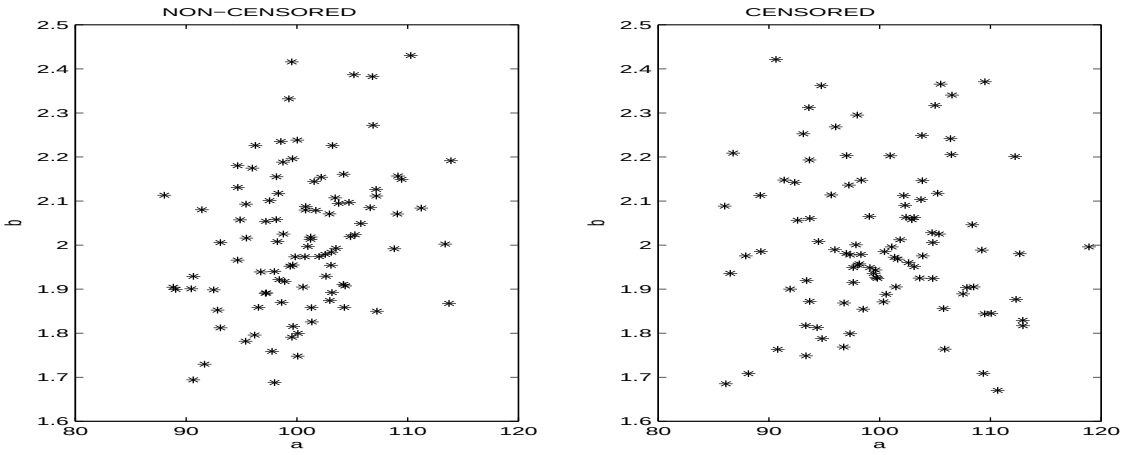


Figure 1: Maximum likelihood estimates of parameters (a_m, b_m) from 100 samples of non-censored (left) and censored data (right)

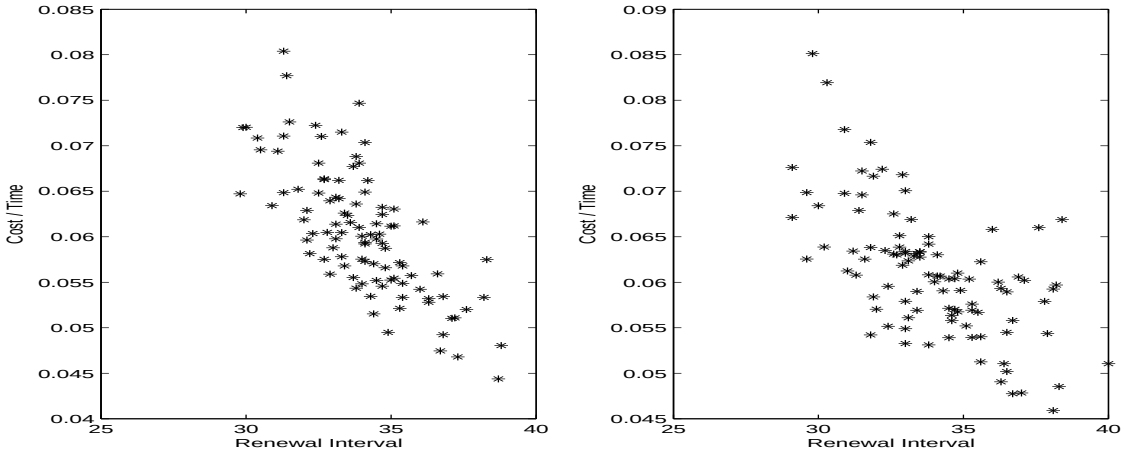


Figure 2: Optimal solutions $(\tau_m, \phi(\tau_m))$ based on estimates displayed in Figure 1, for non-censored (left) and censored (right) cases

Figure 4 displays optimal solutions τ_m^* and $\phi(\tau_m^*)$, each obtained as the solution of (5) with m-th estimate of F, $m = 1, 2, ..., 100$. Again, the left subplot shows the case without censoring, the right subplot then results from censored samples. Notice (expected) larger variability (i.e. uncertainty) in censored data cases.

5 Conclusion

We have studied the impact of variability of statistical estimates to solutions in stochastic optimization problems. We have compared theoretical as well as empirical behavior of estimates in various situations, namely in the parametric or non-parametric setting, cases of fully observed or randomly right-censored data. Influence of variability of estimates to imprecision of optimal solutions has been studied on an example with a model case and randomly generated data.

Acknowledgements

The research is supported by the project of GA CR No 402/10/0956.

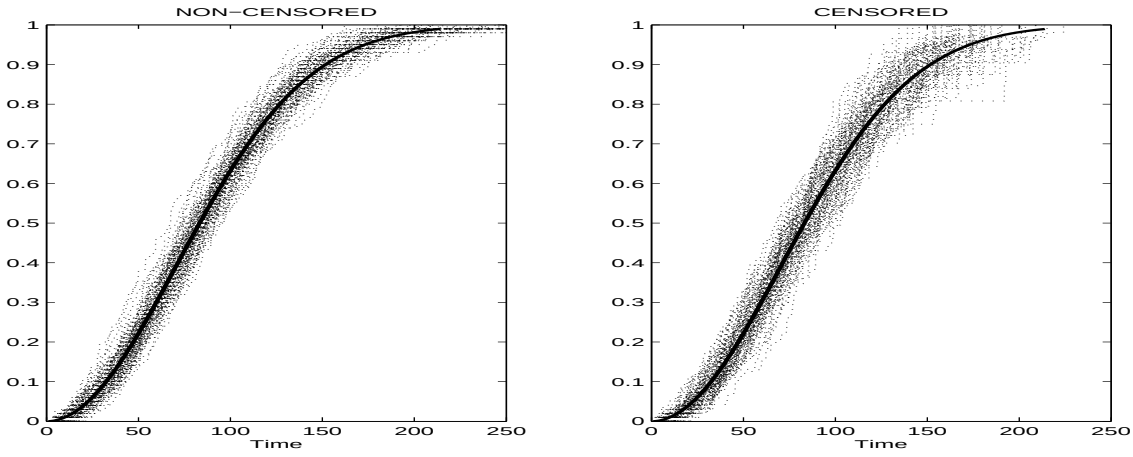


Figure 3: Set of 100 estimates of distribution function, $F_m(t)$, from non-censored (left) and censored data (right). 'True' distribution function $F(t)$ is plotted by solid curve

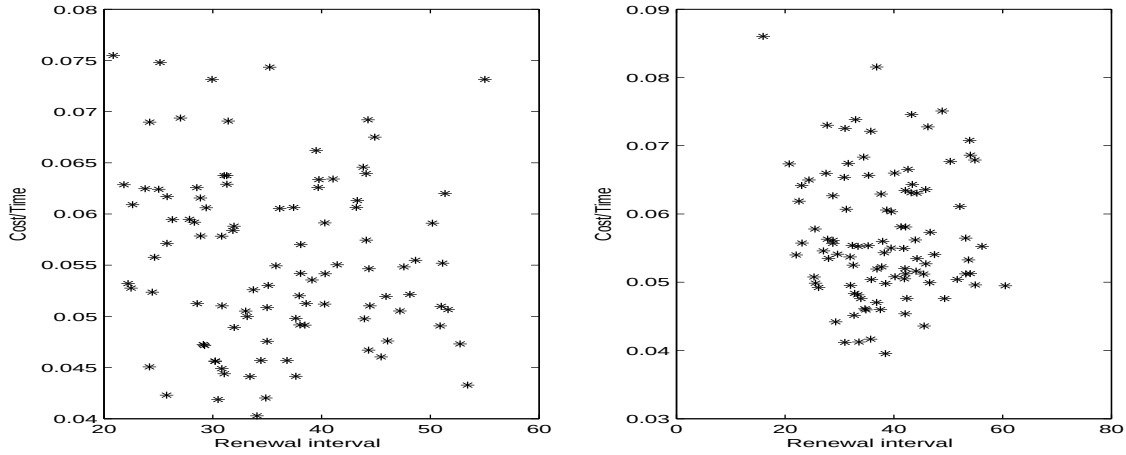


Figure 4: Optimal solutions $(\tau_m, \phi(\tau_m))$ based on nonparametric estimates displayed in Figure 3, for non-censored (left) and censored (right) cases

References

[1] Breslow, N. and Crowley, J. E.: A large sample study of the life table and product limit estimates under random censorship, *Ann. Statist.* **2** (1974), 437–453.

[2] Dupačová, J. and Wets, R.: Asymptotic behaviour of statistical estimates and optimal solutions of stochastic optimization problems, *Ann. Statist.* **16** (1984), 1517–1549.

[3] Huber, C.: Censored and truncated lifetime data. In *Recent Advances in Reliability Theory (N. Limnios and M. Nikulin, eds.)*, Birkhauser, Boston, 2000, 291–305.

[4] Kalbfleisch, J. D. and Prentice, R. L.: *The Statistical Analysis of Failure Time Data (2-nd Edition)*, Wiley, New York, 2002.

[5] Kaňková, V.: A remark of empirical estimates via economic problems. In *Proceedings of MME 2009 (H. Brožová ed.)*, Czech Univ. of Life Sciences, Prague, 2009.

[6] Robbins, H. and Siegmund, D.: Boundary crossing probabilities for a Wiener process and sample sums, *Ann. Math. Statist.* **41** (1970), 14101429.

Models of unexpected fluctuations of aggregate income or real interest rate

Barbora Volná (Kaličinská) [1]

Abstract. In this paper, we create two models based on new IS-LM model: the model of unexpected fluctuations of aggregate income and the model of unexpected fluctuations of real interest rate. The new IS-LM model eliminates two main deficiencies of the original model: assumptions of constant price level and of strictly exogenous money supply. The unexpected fluctuations of these quantities can be explained by existence of special type of cycle called relaxation oscillation. Relaxation oscillations include some short part looking like "jumps". These "jumps" can be interpreted like unexpected. The relaxation oscillation is caused by the fiscal or monetary policy. So, these models with relaxation oscillations can be first approximation of the estimation of the government intervention impacts.

Keywords: new IS-LM model, dynamical behaviour, relaxation oscillations, interest rate, aggregate income.

JEL classification: E12
AMS classification: 37N40, 91B50, 91B55

1 Introduction

In this paper, we create the model of unexpected fluctuations of aggregate income and of real interest rate based on new version of IS-LM model which eliminates two main deficiencies of original IS-LM model.

The original IS-LM [4] has two main deficiencies: an assumption of constant price level and of strictly exogenous money supply, i.e. supply of money is certain constant money stock determined by central bank. There exist many versions of IS-LM model and related problems, see e. g. [1, 2, 3, 6, 7, 9, 10], but we created our oun new version of IS-LM model which eliminates these deficiencies.

The model of unexpected fluctuations of aggregate income is based on relaxation oscillations emerging on goods market under certain conditions and the model of unexpected fluctuations of real interest rate is based on relaxation oscillations on money market or financial assets market under certain conditions. Relaxation oscillations include some short part looking like "jumps". These "jumps" can be interpreted like unexpected. The relaxation oscillations emerging on goods or money market are caused by the fiscal or monetary policy. So, this new models with relaxation oscillations can be first approximation of the estimation of the government intervention impacts.

2 Preliminaries

In this section, there are basic notations and a definition of new IS-LM model which eliminates mentioned two main deficiencies of original IS-LM model.

2.1 Basic Notations

Y	aggregate income (GDP, GNP)	I	investments	L	demand for money
R	long-term real interest rate	S	savings	M	supply of money
i_S	short-term nominal interest rate				

[1] Silesian University in Opava/Mathematical Institute, Department of Real Analysis and Dynamical Systems, Na Rybnicku 1, 746 01 Opava, Czech Republic, e-mail: Barbora.Volna@math.slu.cz

2.2 New IS-LM model

We assume a two-sector economy, a demand-oriented model, $Y \geq 0,\ R \in \mathbb{R},\ i_S \in \mathbb{R}^+$, a variable price level, a conjunction of endogenous and exogenous money supply.

So, we add to original IS-LM model a floating price level, i.e. inflation effect. So, we need to distinguish two type of interest rate. There is the long-term real interest rate on goods market and the short-term nominal interest rate on money (or financial assets) market. The well-known relation holds:

$$i_S = R - MP + \pi^e, \tag{1}$$

where MP is a maturity premium and π^e is an inflation rate. While MP and π^e are constants, it holds $\frac{di_S}{dt} = \frac{d(R-MP+\pi^e)}{dt} = \frac{dR}{dt}$.

Then, we consider that the money supply is not strictly exogenous quantity, but the supply of money is endogenous quantity (money are generated in economics by credit creation, see e.g. [8]) with some exogenous part (certain money stock determined by Central bank).

Definition 2.1. We define the *supply of money* by the formula

$$M(Y, i_S) + M_S, \tag{2}$$

where $M(Y, i_S)$ represents the endogenous part of money supply and $M_S > 0$ represents the exogenous part of money supply.

So, there are the investment function $I(Y, R)$ and the saving function $S(Y, R)$ on goods market and the demand for money function $L(Y, i_S) = L(Y, R - MP + \pi^e)$ and money supply function $M(Y, i_S) = M(Y, R - MP + \pi^e)$ on the money market. It also holds $\frac{\partial L(Y,i_S)}{\partial i_S} = \frac{\partial L(Y,R-MP+\pi^e)}{\partial R}$ and $\frac{\partial M(Y,i_S)}{\partial i_S} = \frac{\partial M(Y,R-MP+\pi^e)}{\partial R}$ because of constant MP and π^e. We suppose that all of these functions are continuous and differentiable.

Definition 2.2. The *new IS-LM model* is given by the following system of two algebraic equations

$$\begin{aligned} &\text{IS:} \quad I(Y, R) = S(Y, R) \\ &\text{LM:} \quad L(Y, R - MP + \pi^e) = M(Y, R - MP + \pi^e) + M_S, \end{aligned} \tag{3}$$

where $M_S > 0$, in the static form and by this system of two differential equations

$$\begin{aligned} &\text{IS:} \quad \tfrac{dY}{dt} = \alpha[I(Y, R) - S(Y, R)] \\ &\text{LM:} \quad \tfrac{dR}{dt} = \beta[L(Y, R - MP + \pi^e) - M(Y, R - MP + \pi^e) - M_S], \end{aligned} \tag{4}$$

where $\alpha, \beta > 0$, in the dynamic form.

Economic theory puts on the main economic functions of this model some properties. We can present these properties using the following formulas:

$$0 < \frac{\partial I}{\partial Y} < 1, \frac{\partial I}{\partial R} < 0, 0 < \frac{\partial S}{\partial Y} < 1, \frac{\partial S}{\partial R} > 0, \tag{5}$$

$$\frac{\partial L}{\partial Y} > 0, \frac{\partial L}{\partial R} < 0. \tag{6}$$

We have to put on our new function of money supply some economic properties. These properties are

$$0 < \frac{\partial M}{\partial Y} < \frac{\partial L}{\partial Y}, \frac{\partial M}{\partial R} > 0. \tag{7}$$

The first formula means that the relation between supply of money and aggregate income is positive and that the rate of increase of money supply depending on aggregate income is smaller than the rate of increase of money demand depending on aggregate income because the banks are more cautious than another subjects. And the second formula means that the relation between supply of money and interest rate is positive.

It is easy to see (using the Implicit Function Theorem) that the course of the curve LM is increasing in this new IS-LM model with properties (6) and (7). If the inequality

$$\frac{\partial I}{\partial Y} < \frac{\partial S}{\partial Y}, \tag{8}$$

also holds in additions to the properties (5), then the course of the curve IS will be decreasing. Now, we denote the function, whose graph is the curve IS, as $R_{IS}(Y)$ and the function, whose graph is the curve LM, as $R_{LM}(Y)$. These functions exist because of the Implicit Function Theorem. Now, if we assume

$$\lim_{Y \to 0^+} R_{IS}(Y) > \lim_{Y \to 0^+} R_{LM}(Y), \tag{9}$$

then there will exist at least one intersection point of the curve IS and LM.

3 Model of unexpected fluctuations of aggregate income (relaxation oscillations on goods market)

In this case, we assume that the subjects and their reactions on goods market are faster than on money market. So, we suppose that the interest rate R is changing very slowly in time in proportion to the aggregate income Y. We can describe this situation by following equations

$$\begin{array}{rcl} \frac{dY}{dt} & = & \alpha[I(Y,R) - S(Y,R)] \\ \frac{dR}{dt} & = & \epsilon\beta[L(Y, R - MP + \pi^e) - M(Y, R - MP + \pi^e) - M_S] \end{array} \tag{10}$$

where $\epsilon \to 0$ is some very small positive parameter.

If this parameter ϵ is very small, then we can consider $\frac{dR}{dt} = 0$ and we can write previous equations in the following forms.

$$\begin{array}{rcl} \frac{dY}{dt} & = & \alpha[I(Y,R) - S(Y,R)] \\ \frac{dR}{dt} & = & 0 \end{array} \tag{11}$$

Now, we formulate sufficient conditions for existence of relaxation oscillations on goods market. We tend to Kaldor's theory about investment and saving function, see [5]. The investment and saving function depends only on Y for some fixed R ($I(Y)$ and $S(Y)$) has so-called "sigma-shaped" graphs, see Figure 1.

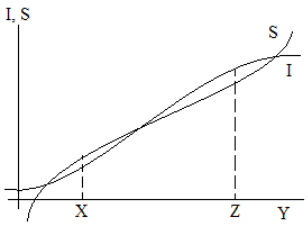


Figure 1: Shapes of the graphs of $I(Y)$ and $S(Y)$

Now, we use (with another designations) the description of these properties from Baráková, see [1]. So-called *"Kaldor's" conditions* are

$$\begin{array}{lll} \frac{\partial I}{\partial Y} < \frac{\partial S}{\partial Y} & \text{for} & Y \in [0, X), \\ \frac{\partial I}{\partial Y} > \frac{\partial S}{\partial Y} & \text{for} & Y \in (X, Z), \\ \frac{\partial I}{\partial Y} < \frac{\partial S}{\partial Y} & \text{for} & Y \in (Z, \infty), \end{array} \tag{12}$$

where points $X < Z$ are given by equation $\frac{\partial I}{\partial Y} = \frac{\partial S}{\partial Y}$ for some fixed R.

Now, we consider the system (10) with requirement of $\epsilon \to 0$ and with properties (5), (6), (7) and (12) (but we can reconsider the system (10) in the simplified way (11)). There we can see that the variable R is a parameter in equations $\frac{dY}{dt} = \alpha[I(Y,R) - S(Y,R)]$. There, every points of the curve IS are singular points. On the next Figure 2 we can see some displaying of relaxation oscillations on goods market.

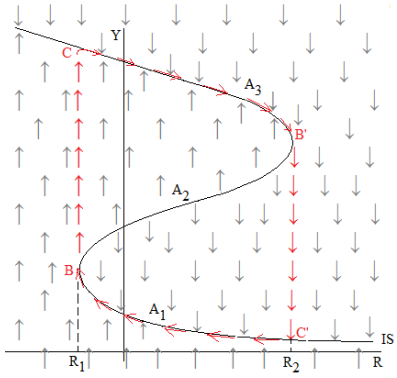


Figure 2: Relaxation oscillations on goods market

The trajectories of this system are directed almost vertically downwards or upwards (parallel to axis Y) considering $\frac{dR}{dt} = 0$. Up or down direction of the trajectories is given by the sign of the function $\beta[L(Y, R - MP + \pi^e) - M(Y, R - MP + \pi^e) - M_S]$ on the curve IS. So, the direction of these trajectories is dependent on the stability or unstability of the arcs A_1, A_2 and A_3. The arcs A_1 and A_3 are stable arcs and the arc A_2 is unstable arc, see Figure 2. It follows from qualification of possible singular points in the system (10) excluding $\epsilon \to 0$. The trajectories are attracted to the stable arcs A_1 or A_3 and are drove away the unstable arc A_2. The velocity of trajectories are finite near the isocline IS and nearness of the curve IS the trajectories go along the curve IS. The velocity of trajectories are infinite large elsewhere. Now, we construct the cycle which is one vibration of the relaxation oscillations. We are changing the parameter R from the level R_2 to R_1. If the moving point is on or near the stable arc A_1, the moving point will go along this stable arc A_1, then it will pass the unstable arc A_2 from point B to C, see Figure 2. The velocity between the point B and C is infinite large. There originates some "jump". There is the similar situation if we are changing the parameter R from the level R_1 to R_2. This oscillation contains the trajectories described by stable arcs A_1 and A_3 with finite velocity and the trajectories described by vertical segments (between points B and C and also between B' and C') with infinite velocity (looking like a "jump"). These trajectories form clockwise cycle.

These changes of the variable R can be caused by the monetary policy and then there originates describing quick "jump". This jump seems to be unexpected.

4 Model of unexpected fluctuations of real interest rate (relaxation oscillations on money market, or financial assets market)

In this case, we assume that the subjects and their reactions on money market are faster than on goods market. So, we suppose that aggregate income Y is changing very slowly in time in proportion to the interest rate R. We can describe this situation by following equations

$$\begin{aligned} \frac{dY}{dt} &= \epsilon\alpha[I(Y,R) - S(Y,R)] \\ \frac{dR}{dt} &= \beta[L(Y, R - MP + \pi^e) - M(Y, R - MP + \pi^e) - M_S] \end{aligned} \tag{13}$$

where $\epsilon \to 0$ is some very small positive parameter.

If this parameter ϵ is very small, then we can consider $\frac{dY}{dt} = 0$ and we can write previous equations in the following forms.

$$\begin{aligned} \frac{dY}{dt} &= 0 \\ \frac{dR}{dt} &= \beta[L(Y, R - MP + \pi^e) - M(Y, R - MP + \pi^e) - M_S] \end{aligned} \tag{14}$$

Now, we formulate sufficient conditions for existence of relaxation oscillations on money or financial assets market. We suppose some unusual behaviour of the demand for money and supply of money. We assume so-called *three phases money demand and money supply* depending on i_S for some fixed Y. In the first phase, for $i_S \in [0, P), P > 0$, these subjects on the money (or financial assets) market behave usual and the course of the money demand and money supply function is standard how we describe in

the condition (6) and (7). In the second phase, for $i_S \in (P, Q), P < Q$, these subjects behave unusual, precisely reversely. We can describe this behaviour using following formula

$$\frac{\partial L}{\partial i_S} = \frac{\partial L}{\partial R} > 0, \frac{\partial M}{\partial i_S} = \frac{\partial M}{\partial R} < 0. \tag{15}$$

These properties correspond to unusual economic situation called liquidity trap. This means that the subjects on money or financial assets market (demand side) prefer liquidity despite relatively high level of (short-term nominal) interest rate. They own money rather than stocks, although they could have bigger gain because of relative high level of (short-term nominal) interest rate. This "irrational" behaviour of these subjects can be caused by big risk of holding these stocks and by small willingness to undergo this risk. The supply of money fully adapts to money demand (we assume demand-oriented model). This phase should be small. In the third phase, for $i_S \in (Q, \infty)$, these subjects behave usual as in the first phase. We can see the graphs of money demand and money supply function describing this behaviour on the following Figure 3.

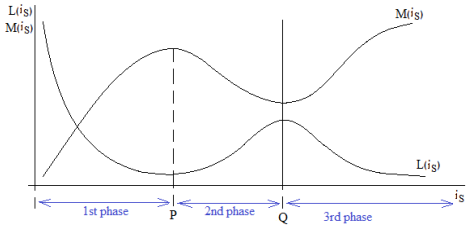


Figure 3: Three phases of the graphs of $L(i_S)$ and $M(i_S)$

Remark 4.1. $\frac{\partial L}{\partial i_S} = \frac{\partial L}{\partial R} = 0$ and $\frac{\partial M}{\partial i_S} = \frac{\partial M}{\partial R} = 0$ in the point P and Q.

Now, we consider the system (13) with requirement of $\epsilon \to 0$ and with properties (5), (8) and three phases money demand and money supply (but we can reconsider the system (13) in the simplified way (14)). There we can see that the variable Y is a parameter in equations $\frac{dR}{dt} = \beta[L(Y, R - MP + \pi^e) - M(Y, R - MP + \pi^e) - M_S]$. There, every points of the curve LM are singular points. On the next Figure 4 we can see some displaying of relaxation oscillations on money or financial assets market.

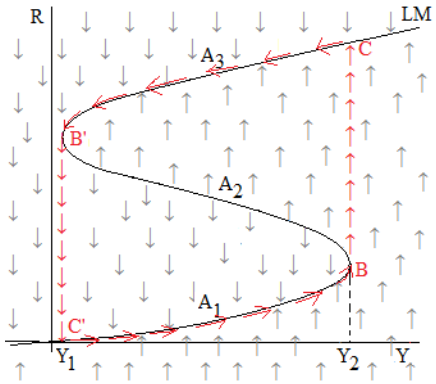


Figure 4: Relaxation oscillations on money (or financial assets) market

The trajectories of this system are directed almost vertically downwards or upwards (parallel to axis R) considering $\frac{dY}{dt} = 0$. Up or down direction of the trajectories is given by the sign of the function $\alpha[I(Y, R) - S(Y, R)]$ on the curve LM. So, the direction of these trajectories is dependent on the stability or unstability of the arcs A_1, A_2 and A_3. The arcs A_1 and A_3 are stable arcs and the arc A_2 is unstable arc, see Figure 4. It follows from qualification of possible singular points in the system (13) excluding $\epsilon \to 0$. The trajectories are attracted to the stable arcs A_1 or A_3 and are drove away the unstable arc A_2. The velocity of trajectories are finite near the isocline LM and nearness of the curve LM the trajectories go along the curve LM. The velocity of trajectories are infinite large elsewhere. Now, we construct the cycle which is one vibration of the relaxation oscillations. We are changing the parameter Y from the level Y_1 to Y_2. If the moving point is on or near the stable arc A_1, the moving point will go along

this stable arc A_1, then it will pass the unstable arc A_2 from point B to C, see Figure 2. The velocity between the point B and C is infinite large. There originates some "jump". There is the similar situation if we are changing the parameter Y from the level Y_2 to Y_1. This oscillation contains the trajectories described by stable arcs A_1 and A_3 with finite velocity and the trajectories described by vertical segments (between points B and C and also between B' and C') with infinite velocity (looking like a "jump"). These trajectories form counterclockwise cycle.

These changes of the variable Y can be caused by the fiscal policy and then there originates describing quick "jump". This jump seems to be unexpected.

Conclusion

In these days, many experts and also the public more and more talk about unexpected fluctuation of different phenomenons in economics and about impact of these fluctuations on economics. We try to model some unexpected fluctuation of aggregate income on goods market and of interest rate on money or financial assets market using own new IS-LM model and theory of relaxation oscillations. New IS-LM model differs from the original model in elimination of its two main deficiencies (assumption of constant price level and of strictly exogenous money supply).

Relaxation oscillations on goods market cause quick change of the aggregate income which seems to be unexpected. Similarly, relaxation oscillations on money market (or financial assets market) cause quick change of the interest rate which seems to be unexpected. This quick "jumps" can ascribe unusual behaviour of economics. Relaxation oscillations on money or financial assets market can be caused by fiscal policy and on goods market by monetary policy. So, this new IS-LM model with relaxation oscillations can be first approximation of the estimation of the government intervention impacts.

Acknowledgements

The research was supported, in part, by the Student Grant Competition of Silesian University in Opava, grant no. SGS/19/2010.

References

[1] BARÁKOVÁ, L.: *Asymptotic behaviour of solutions of differential equations and its applications in economics*, dissertation thesis, Masaryk university, Brno, 2004.

[2] CESARE L. DE; SPORTELLI M.: *A Dynamic IS-LM Model with Delayed Taxation Revenues*, Chaos, Solitons & Fractals, 2005, volume 25, p. 233-244.

[3] GANDOLFO, G.: *Economic Dynamics*, 3rd ed., Springer-Verlag, Berlin-Heidelberg, 1997.

[4] HICKS, J. R.: *Mr. Keynes and the Classics - A Suggested Interpretation*, Econometrica, 1937, volume 5, April, p. 147-159.

[5] KALDOR, N.: *A Model of the Trade Cycle*, Economic journal, 1940, volume 50, March, p. 78-92.

[6] KING, R. G.: *The New IS-LM Model: Language, Logic, and Limits*, Economic Quarterly, Federal Reserve Bank of Richmond, 2000, volume 86/3, Summer, p. 45 - 103.

[7] KODERA, J.: *Non-linear Dynamics in IS-LM Model*, Proceedings of the International Conference Mathematical Methods in Economics, University of West Bohemia, Cheb, 1999, p. 59 - 64.

[8] SOJKA, M.: *Monetární politika evropské centrální banky a její teoretická východiska pohledem postkeynesovské ekonomie (Monetary Policy of the European Central Bank and Its Theoretical Resources in the View of Postkeynesian Economy)*, Politická ekonomie, 2010, issue 1, p. 3 - 19.

[9] VARIAN, H. R.: *The Stability a Disequilibrium IS-LM Model*, The Scandinavian Journal of Economics, 1977, volume 79, No. 2, p. 260 - 270.

[10] ZHOU, L.; LI, Y.: *A Generalized Dynamic IS-LM Model with Delayed Time in Investment Process*, Applied Mathematics and Computation, 2008, volume 196, Issue 2, March, p. 774 - 781.

Modeling of claim counts using data mining procedures in R CRAN

Alicja Wolny-Dominiak[1]

Abstract. In the ratemaking process the ranking, which takes into account the number of claims generated by a policy in a given period of insurance, may be helpful. For example, such a ranking allows to classify the newly concluded insurance policy to the appropriate tariffs group. For this purpose, in this paper we analyze models applicable to the modelling of counter variables. In the first part of the paper we present the classical Poisson regression and a modified regression model for data, where there is a large number of zeros in the values of the counter variable, which is a common situation in the insurance data. In the second part we expand the classical Poisson regression by adding the random effect. The goal is to avoid an unrealistic assumption that in every class all insurance policies are characterized by the same expected number of claims. In the last part of the paper we propose to use k-fold cross-validation to identify the factors which influence the number of insurance claims the most. Then, setting the parameters of the Poisson distribution, we create the ranking of polices using estimated parameters of the model, which give the smallest cross-validation mean squared error and we classify using the regression tree. In the paper we use a real-world data set taken from literature. For all computations we used a free software environment R.

Keywords: claim counts, ZIP, HGLM, R CRAN

JEL Classification: C15, C88
AMS Classification: 65C60

1 Introduction

Every person, when applying for an insurance policy, is assigned to a class, that is homogeneous in terms of the rate-making process. One of the criteria used for assigning an individual to a certain class is the number of claims. Thus it is insurance companies' very important task to model the number of claims in a given insurance portfolio. In the paper we propose a simple procedure for creating a ranking of insurance policies and also for classifying them due to the number of claims. It allows a preliminary classification of a new policy to a group with an adequate premium level.

The very common choice of a method for modelling the number of claims is a regression model as in [1] with the use of Poisson distribution, which is a special case of a Generalized Linear Model (GLM Poisson), see in [8]. In regression claims modelling, dependent variables may be interpreted as risk factors. For the selection of these variables into the model one may use traditional methods from [9] or adopt genetic algorithms as in [3]. However the insurance portfolios have a very specific characteristic, i.e. for many policies there are no claims observed in the insurance history for a given period. It means that the data contains lots of zeros and, as a consequence, the Poisson regression may not give satisfactory results which is shown in [11]. Also these two models are with fixed effects so the assumption of independence among the responses is necessary which is sometimes unrealistic for insurance data set. To avoid this problems, except GLM Poisson model we considered ZIP Poisson model according to [6] and HGLM Poisson – Gamma model with a random effect as in [7].

The ranking creation procedure used a *k*–fold cross–validation and furthermore the ranking was discretized due to a parameter λ. We build many different models and then we use a 10–fold cross–validation in order to recognize which rating variables have an impact on the presence of zeros in the policies portfolios. Finally in order to simplify the ranking and classification we applied a regression tree. The data for the illustrative example has been taken from the literature [10]. All the computations were conducted in R – the free software environment. The procedure for building a model with random effect and a cross–validation technique have been written in R language.

[1] University of Economic in Katowice, Department of Statistical and Mathematical Methods in Economics, alicja.wolny-dominiak@ue.katowice.pl.

2 Modelling the number of claims

The generalized linear models (GLM) are used for creating a raking of insurance policies due to the number of claims. In GLM we assume that the number of claims is a dependent variable Y that follows a Poisson distribution and it depends on a certain system of predictors as in [2]:

$$P(Y_i = y_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!},\ i = 1,...,n$$

where Y_i is the number of claims for the *i*–th insured person, $y_i,..., y_n$ are independent and have equal variances, and the average number of claims is equal to the variance. The λ_i parameter is the expected number of claims and it depends on predictors X_j, $j = 1,...,k$ that describe the insured individual or vehicle, e.g. sex, age, engine capacity. The logarithm is used as a link function as follows:

$$\ln \lambda_i = \sum_{j=1}^{k} \beta_{ji} X_{ji}$$

When creating the ranking we used $\min \lambda_i$ as a criterion.

The independence assumption in the above model may not be fulfilled. In that case the solution is to use a HGLM Poisson-Gamma and introducing a random effect v as in [5]. In case of automobile insurance data "Region" or "Vehicle model" can be treated as a random effect v. The HGLM Poisson-Gamma model has a form according to [7]:

$$\mu = E(y \mid u) = e^{X\beta + v}$$
$$v = \log u$$

where $\beta = [\beta_1,..., \beta_I]$, $u = [u_1,..., u_K]$ and **X** is the model matrix. The structural parameters of a model have a following interpretation:

- parameter β_i, $i = 1,..., I$, measures the influence of the *i*–th predictor on the number of claims;
- parameter u_k, $k = 1,..., K$, measures the risk level for every category (which is different for every category).

Another model used for modelling the number of claims is ZIP model, where counting response variable has a lot of zero values. This is exactly the case when modelling the number of counts. Analysing different risk portfolios it can be noticed that for many policies there is no claim observed and if the claims occur their number is one, two or three and very rarely more. In the ZIP model the independent variables Y_i take zero values $Y_i \sim 0$ with the probability ϖ_i or values from Poisson distribution $Y_i \sim Pois(\lambda_i)$ with probability $1-\varpi_i$. It can be written in a form as in [6]:

$$P(Y_i = y_i) = \begin{cases} \varpi_i + (1-\varpi_i)e^{-\lambda_i}, y_i = 0 \\ (1-\varpi_i)\dfrac{e^{(-\lambda_i)\lambda_i^{y_i}}}{y_i!}, y_i > 0 \end{cases},\ i = 1,..,n$$

Thus in the ZIP model we have two parameters: λ_i and ϖ_i. Both parameters, as in case of Poisson regression, are linked with predictor variables with the following link functions:

$$\ln\left(\frac{\varpi_i}{1-\varpi_i}\right) = \sum_{j=1}^{t} \gamma_{ji} Z_{ji}$$
$$\ln \lambda_i = \sum_{j=1}^{k} \beta_{ji} X_{ji},$$

where $Z_1,..., Z_l$ are the dependent variables for the first equation and $X_1,..., X_k$ for the second one. Similarly to Poisson regression case, in the ZIP model we assume that the average number of claims equals the variance. The solution to a problem when overdispersion occurs is the use of negative-bimodal distribution (ZINB model), see in [6].

3 Procedure of creating ranking of property insurance policies and classification of these policies

The procedure of building a ranking of policies using linear models presented in the previous part of the paper may be formulated in a few steps:

1. Estimating λ parameter for every policy in the portfolio using 3 different models: GLM, HGLM and ZIP/ZINB model;

2. Applying 10–fold cross–validation procedure to every model from Step 1 [4]:
   a) randomly divide the training set into $k = 10$ approximately equally sized parts (n - the training set size, m_l - the size of the l -th subset, $l = 1,...,10$),
   b) build 10 times a model using 9 of 10 parts ($n - m_l$ observations), treating excluded observations as validation set,
   c) calculate 10 times the value of the mean squared error $MSE_i = \frac{\sum (y - \mu_l)^2}{m_l}$ using the validation set,
   d) estimate the cross–validation error: $cv = \sum_{l=1}^{10} \frac{m_l}{n} MSE_l$. The model with the smallest cv value is selected,
3. Choosing the model with the smallest cv error;
4. Creating the ranking of insurance policies for every combination of predictor variables X_i , using as a criterion $MIN\lambda$;
5. Discretizing the ranking due to the values of parameters λ and thus obtaining insurance risk classification which allow to classify a new policy to a group with an adequate premium level.

Based on the estimated parameter λ for a chosen model, we have created ranking and conducted discretization in order to obtain different classes of insurance risk. Discretization means dividing the ordered set of values of a given continuous variable onto finite number of disjoint intervals. Labels can be assigned to these intervals, e.g. high insurance risk level, neutral to risk etc. The problem is how to determine the cut points. These cut points should separate the object from different risk classes in a best possible way. There are two main approaches in discretization: agglomerative and divisive. The first one starts with every single empirical value of the continuous variable belonging to a different interval and then neighbouring intervals are merged iteratively until the maximum value of a homogeneity of subsets measure is reached. The second approach starts with one big interval covering all empirical values of the continuous variable and then it is iteratively divided, using previously determined cut points.

4 Case study for automobile insurance data set

In order to illustrate the process of creating the ranking and discretizing it, the necessary procedures were implemented in R environment. The automobile insurance data set including information about the number of claims has been used for computations [10]. The following variables form the data set and have been considered in the model:

1. *Driver.age* – age of the insured person (driver);
2. *Region*: classes from 1 to 7;
3. *MC.class*: classes from 1 to 7 which were created based on the EV coefficient defined as $EV = \frac{\text{engine capacity in kW x 100}}{\text{vehicle weight in kg} + 75}$, where 75 kg is the average weight of a driver;
4. *Veh.age* – age of the vehicle;
5. *Num.claims* – number of claims – the sum within the class.

Procedure for creating the ranking

1. We model the number of claims with the use of three types of models presented above.

Model 1. GLM for the variable *Num_claims* assuming Poisson distribution

R Code

```
data(dataset)
glm.formula=Num.claims~Driver.age+Region+MC.class+Veh.age
glm.model1=glm(glm.formula, family=Poisson(link="log"), data=dataset)
summary(glm.model1)
```

Model 2. HGLM of a type POISSON-GAMMA for the variable *Num.claims* assuming Poisson distribution and treating variable *Region* as a random effect with Gamma distribution

R Code

```
Library(hglm)
data(dataset)
```

```
hglm.model2=hglm(fixed= Num.claims~Driver.age+Region+MC.class+Veh.age, random=~1|Region, family=Poisson(link="log"), rand.family = family=Gamma(link="log"), data=dataset)
summary(hglm.model2)
```

Model 3. Model ZIP taking into account a large number of zero values for variable *Num_claims*
R Code

```
Library(pscl)
data(dataset)
ZIP.model3=zeroinfl(formula=Num.claims~Driver.age+Region+
MC.class+Veh.age| Driver.age+Region+MC.class+Veh.age, data=dataset)
summary(ZIP.model3)
```

Function `zeroinfl` is from the library `{pscl}`

2. Ten fold cross–validation procedure was applied to every model from Step 1, obtaining corresponding *cv* errors which are shown in Table 1.

Model	cross-validation error
GLM Poisson	10.76
HGLM Poisson-Gamma	2.15
ZIP	0.89

Table 1 Cross-validation errors

3. The smallest value of *MSE cv* was obtained for the Model 3., i.e. for the zero–inflated generalized linear model. Thus this model was used further in the ranking creation steps. The results are presented in Table 2.

variables	parameters	standard error	tariffs
Intercept	-1.179	0.303	0.308
Driver_ageA	0.000	-	1.000
Driver_ageB	-0.269	0.189	0.764
Driver_ageC	-0.514	0.189	0.598
Driver_ageD	-1.281	0.202	0.278
Driver_ageE	-1.305	0.187	0.271
Driver_ageF	-1.447	0.198	0.235
Driver_ageG	-1.976	0.296	0.139
RegionA	0.000	-	1.000
RegionB	-0.385	0.112	0.681
RegionC	-0.807	0.121	0.446
RegionD	-0.898	0.108	0.407
RegionE	-1.831	0.345	0.160
RegionF	-1.446	0.251	0.235
RegionG	-2.048	1.011	0.129
MC_classA	0.000	-	1.000
MC_classB	0.320	0.204	1.377
MC_classC	0.081	0.171	1.084
MC_classD	-0.007	0.183	0.993
MC_classE	0.560	0.174	1.751
MC_classF	1.046	0.172	2.846
MC_classG	-0.479	0.444	0.619
Veh_ageA	0.000	-	1.000
Veh_ageB	-0.459	0.127	0.632
Veh_ageC	-0.771	0.129	0.463
Veh_ageD	-1.241	0.112	0.289

Table 2 Parameters for Model 3 - ZIP

The probability that variable *Num.claims* takes zero value equals 82%.

4. After discretization every combination was assigned a label representing a risk class: from A – the lowest risk of claim to occur, to J – the highest risk of claim to occur. The distribution of risk classes is as follows in Table3.:

risk class	number of policies in classes	% of policies in classes
A	949	69.17%
B	241	17.57%
C	100	7.29%
D	45	3.28%
E	17	1.24%
F	10	0.73%
G	6	0.44%
H	2	0.15%
J	2	0.15%

Table 3 Distribution for risk classes - Model 3

So finally we received 9 risk classes. The number of combinations of different empirical values of predictor variables X_i equals 1372. In order to obtain a more synthetic description of each risk class, the classification tree model was used. The description of the this tree are presented in Table 4.

class description	class
If (Region=EFG)	A
If [(Region=ABCD) and (MC.class=AG)]	A
If [(Region=ABCDJ) and (MC.class=BCDEGF) and (Drive.age=G)]	A
If [(Region=ABCDJ) and (MC.class=BCDEF) and (Drive.age=GABCDF) and (Veh.age=ABC)]	A
If [(Region=ABCDJ) and (MC.class=AGCDEF) and (Drive.age=EF) and (Veh.age=ABC)]	A
If [(Region=ABCD) and (MC.class=BCDEFAG) and (Drive.age=ABCDEF) and (Veh.age=ABC)]	B
If [(Region=ABCDJ) and (MC.class=F) and (Drive.age=EF) and (Veh.age=ABC)]	B
If [(Region=ABCD) and (MC.class=CDEF) and (Drive.age=BCE) and (Veh.age=ABC)]	B
If [(Region=ABD) and (MC.class=CDEF) and (Drive.age=ABCDFE) and (Veh.age=ABC)]	B
If [(Region=ABD) and (MC.class=BCDEF) and (Drive.age=ABCDEF) and (Veh.age=D)]	B
If [(Region=ABD) and (MC.class=BCDEF) and (Drive.age=ABCDEF) and (Veh.age=D)]	B
If [(Region=ABD) and (MC.class=CDEF) and (Drive.age=BCE) and (Veh.age=ABC)]	C
If [(Region=ABD) and (MC.class=F) and (Drive.age=ABCDF) and (Veh.age=ABC)]	C
If [(Region=ABD) and (MC.class=CDEF) and (Drive.age=ABCDEF) and (Veh.age=D)]	C
If [(Region=ABD) and (MC.class=BCDEF) and (Drive.age=BCE) and (Veh.age=D)]	E

Table 4 Regression tree for risk classes

R Code

```
library(rpart)
library(e1071)
data(dataset)
data.tree=NULL
data.tree=data[,c(1:4, ncol(data))]
model.formula=paste(names(data)[ncol(data)] ," ~ .", sep="")
model.rpart=tune.rpart(eval(parse(text=model.formula)), data=data.tree, minsplit=3:10, cp=c(0.01, 0.03, 0.05))#,
method = "class")
summary(model.rpart)
table(dane[,ncol(data)])
print(model.rpart$best.model)
plot(model.rpart$best.model)
text(model.rpart$best.model)
```

5 Summary

The procedure for recognizing risk classes in the insurance policies portfolios proposed in the paper allows to differentiate policies with no claims observed in the insurance history. The minimum value of λ criterion used in classification causes that the risk classes and associated premiums are fairer for individuals applying for an insurance policy. Essentially the main disadvantage of ZIP model, that turned out to be the best in terms of *cv* error criterion, is that within every risk class the policies have equal expected number of claims, which is an unrealistic assumption. The solution to this issue may be using the mixed Poisson model and introducing a random effect that would differentiate policies (ZIP regression with random effect). However estimating that type of model is computationally very demanding what discourages from using in real world applications.

Acknowledgements

This research is supported by the grant of Polish Ministry of Science and Higher Education (nr NN 111461540).

References

[1] De Jong P., Heller G.Z.: *Generalized Linear Models for Insurance Data*. Cambride University Press, Cambrige, 2008.

[2] Denuit M., Marechal X., Pitrebois S., Walhin J.: *Actuarial Modelling of Claims Counts*. John Wiley&Sons Ltd, 2007.

[3] Gamrot W.: Representative Sample Selection via Random Search with Application to Surveying Communication Lines, *Proceedings of 26th, International Conference on Mathematical Methods in Economics 2008, (Rehorova P., Marsikova K, Hubinka Z., eds.)*, Technical University of Liberec, 127-132.

[4] Gatnar E.: *Ensemble Approach in Classification and Regression* (in Polish). Wydawnictwo Naukowe PWN, Warszawa 2008.

[5] Hall D. B.: Zero-Inflated Poisson and Binomial Regression with Random Effects: A Case Study, *Biometrics* **56** (2000), 1030-1039.

[6] Lambert D.: Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics* 34 (1992), 1-14.

[7] Lee Y., Nelder A. J., Pawitan Y.: *Generalized Linear Models with Random Effects*. Monographs on Statistics and Applied Probability 106, Chapman&Hall\CRC, New York, 2006.

[8] McCullagh P., Nelder J. A.: *Generalized Linear Models*. Chapman & Hall/CRC, New York, 1999.

[9] Miller A.: *Subset selection in Regression*. Chapman and Hall, London, 1990.

[10] Ohlsson E., Johansson B.: *Non-Life Insurance Pricing with Generalized Linear Models*. Springer-Verlag, Berlin, 2010.

[11] Wolny-Dominiak A.: Zero-Inflated Poisson Model for Insurance Data with a Large Number of Zeros (in Polish), In: *Forecasting in Management*, *Research Papers of Wroclaw University* **185** (2011), 21-30.

Insurance portfolios rate making: Quantile regression approach

Alicja Wolny–Dominiak[1], Agnieszka Ornat–Acedańska[2], Grażyna Trzpiot[3]

Abstract. Insurance portfolios rate–making frequently based on different multivariate regression models which is more sensitivity to the assumptions which significantly restrict the area of their applications. When error term is non-normal, asymmetric, fat–tailed or in presence of outliers it may have serious consequences for correct inference on the factors impact on endogenous variable. In this paper we analyze three models: generalized linear model (GLM), hierarchical generalized linear model (HGLM) and exponential quantile regression model (EQRM) in rate-making process. The approach using EQRM is robust to deviations from the classical assumptions (a distribution of error terms is left unspecified, which is the main virtue of the method as far as robustness to outliers is concerned). The aim of this paper is to applied GLM, HGLM and EQRM for rate–making and analyzed real insurance automobile data set. For this data set we adopted cross–validation procedure to compare results of these models according to cross–validation Root MSE criteria.

Keywords: rate–making, quantile regression, hierarchical generalized linear models, cross–validation.

JEL Classification: C21, G22
AMS Classification: 65C60

1 Introduction

The rate–making process is one of the most important problem in insurance portfolios issues. The techniques of rate–making are actually based on loss distribution or their moments, which are estimates using historical data. The key problem is to choose the correct model for estimation of loss value. Insurance portfolios rate–making is frequently based on different multivariate regression models which allow to investigate rating factors. Nevertheless, ordinary multivariate regression model has some crucial disadvantages – it is sensitive to the assumptions which significantly restrict the area of their applications. In insurance data case, when error term is non-normal, asymmetric, fat-tailed or in presence of outliers it may have serious consequences for correct inference on the factors impact on endogenous variable. Moreover, ordinary multivariate regression model often ignores the specific feature of the insurance data used. For example, for the real insurance portfolio, there are: possibility of catastrophic losses, the dependence of insured objects on each other (i.e. cumulating risk) or information shortfall to verify the statistical significance of model chosen [5]. Therefore, for modeling insurance data it is important to use models and estimators that are more robust to restrictive classical regression assumptions.

GLM is a good example of such model and therefore it is used by actuaries [7], [8], [1], [9]. However there are some problems connected with GLM. First problem is in choosing the predictors' distribution in GLM. It can be solve with simulation procedure based on the Monte Carlo method [20]. Second problem is in independency assumption for the value of claims. In such a situation HGLM model is recommended. The other approach proposed for modeling insured data (in particular expected net premium rates) is quantile regression – see Kudryavtsev [5]. This approach is consistent with the idea of using the distribution quantile for rate-making. Additional advantage of this method is fact that it allows to estimate the net premium rates including safety loadings and it may be estimated as a quantile of loss distribution.

The first section contains the description of Generalized Linear Model and Hierarchical Generalized Linear Model for the rate-making. Then we discuss the methodology of the quantile regression including it's special case – the exponential quantile regression model as the model which is used for the rate-making. The next section contains the description of the Cross-validation procedure. The empirical results are presented in section four.

[1] University of Economics in Katowice, Faculty of Economics, Department of Statistical and Mathematical Methods in Economics, Bogucicka 14, 40–226 Katowice, alicja.wolny-dominiak@ue.katowice.pl
[2] University of Economics in Katowice, Faculty of Informatics and Communication, Department of Demography and Economic Statistics, Bogucicka 14, 40–226 Katowice, agnieszka.orwat@ue.katowice.pl
[3] University of Economics in Katowice, Faculty of Informatics and Communication, Department of Demography and Economic Statistics, Bogucicka 14, 40–226 Katowice, trzpiot@ue.katowice.pl

2 Generalized linear model and hierarchical generalized linear model for rate–making

Currently in practice of insurance rate-making GLM models are applied. In GLM, a continuous dependent variable Y_i is treated as the value of claims in portfolio and categorical explanatory variables $X_{i1},....,X_{im}$, $i=1,2,\ldots,n$ are treated as rating factors. Observations $Y_1,...,Y_n$ are assumed to be independent. In insurance data the dependent variable Y_i is usually non-negative and skewed to the right. That is why the Gamma GLM for rate-making is applied in practice and is of the form $\mathbf{Y}=g^{-1}(\mathbf{X\beta})$, where $Y\sim G(\mu,\sigma)$ and the link function $g(x)=\ln x$ [1]. Parameters $\beta_1,...,\beta_m$ measure the impact of rating factors on the value of claims (constant for all categories). Tariff rates are calculated by the formula:

$$t=\exp(\mathbf{X\beta}) \tag{1}$$

and show how to change the base premium calculated as $\exp(Intercept)$ for every category of rating factors.

In practice independent assumption for the variable Y is unrealistic. In this case the linear mixed models are useful, where one categorical explanatory variable is assumed to be a random effect. The example of such a variable is area or vehicle model in third party claims. Then the rate-making process is carried out for every category of random effect separately. In this process the HGLM gamma-gamma model can be applied [19] and is of the form [6]:

- conditional on random effects u, the responses Y_i follow a GLM gamma family, satisfying:

$$\mathbf{Y}=g^{-1}(\mathbf{X\beta}+Zv),\ v(u)=\ln u \tag{2}$$

- the random effect u follows a distribution conjugate to a GLM gamma family

The fixed effects $\beta=[\beta_1,...,\beta_l]$ and random effects $v=[v(u_1),....,v(u_K)]$ of the model have the following interpretation:

- parameters β_i, $i=1,...,m$, measure the impact of i-th rating factor on the value of claims (constant for all categories)
- parameters $v(u_k)$, $k=1,...,K$, measure the level of risk within the category (inconstant for all categories)

Similarly as in GLM gamma, tariff rates are determined by the formula:

$$t=\exp(\mathbf{X\beta})\exp(Zv) \tag{3}$$

and show how to change the base premium calculated as $\exp(Intercept)$ for every category of fixed rating factors adjusted for the tariff cell for random effect.

3 Quantile regression

In quantile regression method [2] we analyze a problem of estimation of a vector of parameters $\boldsymbol{\beta}$ for a sample of independent observations y_i, $i=1,2,\ldots,n$ of a sequence of random variables $Y_1,Y_2,..,Y_n$ taken with distribution $P(Y_i<y)=\Im(y-\mathbf{x}_i'\boldsymbol{\beta})$, where $\mathbf{x}_i=(1,x_{i1},x_{i2},..,x_{im})'$ is a column $(n\times m+1)$-matrix of observations $\mathbf{X}$ and the distribution $\Im$ is unknown. The linear quantile regression model (LQRM) of order τ, $0<\tau<1$ is given by the formula [3]:

$$Q_\tau(Y_i|\mathbf{x}_i)=\mathbf{x}_i'\boldsymbol{\beta}^{(\tau)} \tag{4}$$

where $Q_\tau(Y_i|\mathbf{x}_i)$ indicates conditional quantile of random variable Y_i for probability τ provided vector $\mathbf{x}_i$, $\boldsymbol{\beta}^{(\tau)}=(\beta_0^{(\tau)},\beta_1^{(\tau)},\beta_2^{(\tau)},..,\beta_m^{(\tau)})'$ is vector of regression coefficient. The LQRM corresponding to the to linear regression model (LRM) can be expressed as $Y_i=\beta_0^{(\tau)}+\beta_1^{(\tau)}X_{i1}+\ldots+\beta_{im}^{(\tau)}X_{im}+\varepsilon_i^{(\tau)}$, where $\varepsilon_i^{(\tau)}$ is error term. Then $Q_\tau(\varepsilon_i^{(\tau)}|\mathbf{x}_i)=0$. A distribution of independent random variables $\varepsilon_i^{(\tau)}$ is left unspecified, which is the main virtue of the method as far as robustness to outliers is concerned. Koenker and Basset [2] defined a τ-th quantile regression estimator of $\boldsymbol{\beta}^{(\tau)}$, that its value $\mathbf{b}$ solves the problem:

$$\min_{\mathbf{b}\in\Re^{m+1}}\left[\sum_{i\in\{i:\,y_i\geq x_i'\mathbf{b}\}}\tau\left|y_i-\mathbf{x}_i'\mathbf{b}\right|+\sum_{i\in\{i:\,y_i<x_i'\mathbf{b}\}}(1-\tau)\left|y_i-\mathbf{x}_i'\mathbf{b}\right|\right] \tag{5}$$

The problem (5) has always a solution; for continuous distributions it is unique. Since the problem (5) can be transformed to a linear optimization problem its solution can be find using an internal point method [12].

Because the error distribution term is unspecified, statistical inference is based on nonparametric approach – bootstrap or Monte Carlo method. In bootstrap approach samples are drawn with replacement from analyzed data set. Based on the sample of n observations we form a bootstrap sample drawing *n* observations from the original sample. The procedure is repeated *N* times $(N \geq 1000)$. For every bootstrap sample *k* we calculate estimates $\beta_{kj}^{(\tau)}$. Then for hypothesis testing of parameter significance we calculate fraction of samples for which $\beta_{kj}^{(\tau)} = 0$ (null hypothesis $H_0 : \beta_j^{(\tau)} = 0$, alternative hypothesis $H_1 : \beta_j^{(\tau)} \neq 0$) and treat it as an empirical *p-value* of the test.

The quantile approach detects relationships missed by traditional data analysis methods. Robust estimates detect the influence of the bulk of the data, whereas quantile estimates detect the influence of co-variates on alternative parts of the conditional distribution. In practice, the distribution-free approach is often used for estimation – see for example Koenker, Halllock [4], Koenker [3]. Applications of the quantile regression for Polish capital market can be found in Trzpiot [13], Trzpiot [14], Trzpiot [15], Trzpiot [16], Trzpiot [17] and modifications of quantile regression in Orwat–Acedańska, Trzpiot [10], Orwat–Acedańska, Trzpiot [11] and Trzpiot [18].

3.1 Exponential quantile regression model for rate–making

Taking more general type of model ($\mathbf{Y} = g^{-1}(\mathbf{X}\boldsymbol{\beta})$) then linear models allows an actuary to take into account the influence of risk factors on the loss amount in the framework of linear form while the model is non-linear [5]. Combining this formulae with model (4) one can use the method of quantile regression in the following way $Q_\tau(Y_i|\mathbf{x}_i) = g^{-1}(\mathbf{x}_i'\boldsymbol{\beta}^{(\tau)})$.)]. One of the possibilities is taking logarithm as a function *g*.[4] Then $\mathbf{Y} = \exp(\mathbf{X}\boldsymbol{\beta})$.

In this paper we taking as function $g(\cdot)$ logarithm and we use following the exponential quantile regression model (EQRM) of order τ^*:

$$Q_{\tau^*}(Y_i|\mathbf{x}_i) = \exp(\mathbf{x}_i'\boldsymbol{\beta}^{(\tau^*)}), \tag{6}$$

where $Q_{\tau^*}(Y_i|\mathbf{x}_i)$ indicates conditional quantile of random variable Y_i for probability τ^*, $0 < \tau^* < 1$ provided vector of rating factors $\mathbf{x}_i$, and $\boldsymbol{\beta}^{(\tau^*)} = (\beta_0^{(\tau^*)}, \beta_1^{(\tau^*)}, \beta_2^{(\tau^*)}, \ldots, \beta_m^{(\tau^*)})'$ is vector of regression coefficient of order τ^*.

The exponential quantile approach to making net premium rate is base on probability: [5].

$$\tau^* = \frac{\tau - p}{1 - p}, \text{ where } 0 < \tau^* < 1 \tag{7}$$

where *p* is fraction of policies with no claims incurred.

Model (6) of order τ^* in the form (7) gives the tariff rate estimators that are convenient for practical use: the estimators are conditional quantiles (given observable rating factors $\mathbf{x}_i$ known) of probability τ^* for *i*-th policy before loss occurs [5]. Finally tariff rates are determined by the formula (1).

4 Cross–validation procedure

In order to unify the process of comparing presented models, the choice of the model for rate-making is supported by statistical learning methods. In general in these methods we assume we are given a training data set $D = \{(x^i, y^i), i = 1,..., N\}$, where $x^i, y^i \in R$. Moreover we assume that data is i.i.d. (independent and identically distributed) and it has been taken from the population with a multidimensional distribution defined by an unknown density function:

$$p(x, y) = p(x)p(y \mid x) \tag{8}$$

The task is to search a given set of functions $H = \{f(x,\varpi) : \varpi \in \Omega\}$, where ϖ is a model parameters vector, and to find the best element. Using the model $f(x,\varpi) \in H$, which is always a simplified equivalent of the analysed phenomenon, we accept some errors that are just the consequence of taking theoretical values instead of real values for response variable. These errors (for a given observation) are measured by so called *loss functions* $L(y, f(y,\varpi))$. In the concept of statistical learning the risk functional is considered which measures the overall loss, i.e. the sum of errors for all possible observations. One of the methods of estimating the value of the risk functional is the cross–validation method (*CV*) . This paper uses 5–fold cross–validation algorithm for all models, i.e.:

[4] Function $g(\cdot)$ is usually differentiable and monotonic.

1. randomly divide the training set into $k = 5$ approximately equally sized parts;
   (n – the training set size, m_l – the size of the l – th subset, $l = 1,...,5$);
2. build 5 times every model using 4 of 5 parts ($n - m_l$ observations), treating excluded observations as validation set;
3. calculate 5 times the value of the mean squared error $RMSE_l = \sqrt{\frac{\sum (y - \hat{\mu}_l)^2}{m_l}}$ using the validation set;
4. estimate the cross–validation error: $cv = \sum_{i=1}^{5} \frac{m_l}{n} MSE_l$;

The model with the smallest cv value is selected.

5 Results of empirical analysis

In order to illustrate the process of rate–making with EQRM, GLM and HGLM models, the empirical example was calculated using the automobile insurance data set from literature [9]. The following variables from the data set have been considered in models:

1. `Driver.age` – age of the insured person (driver):
2. `Region`: classes from A to G
3. `MC.class`: classes from A to G

   These classes were created based on the *EV* coefficient defined as:

   $EV = \frac{\text{engine capacity in kW x 100}}{\text{vehicle weight in kg} + 75}$, where 75 kg is the average weight of a driver;
4. `Veh.age` – age of the vehicle.

For the data set GLM, HGLM and EQRM models were applied with following assumptions:

- GLM – all rating factors are fixed effects with Gamma distribution (gamma model)
- HGLM – the risk factor `Region` is the random effect with Gamma distribution (Gamma–Gamma model)
- EQRM – the parameter $\tau = 0.99$, and on the base of empirical data we computed the fraction of policies with "no losses" incurred: p = 0.96. Thus, according to formula (6) we implemented EQRM of order $\tau^* = 0.75$. Moreover, p–values in EQRM were calculated using bootstrap method.

For GLM and HGLM models estimation we used a free software environment R CRAN, the package `{stats}` (`glm`) and the package `hglm` (function `hglm`). In the case EQRM we our own procedures, which were created in Matlab program in order to parameters estimate and to test of parameter significance. The *p-values* were calculated by means of bootstrap method. The estimated tariff rates in analyzed threes models are in table 1.

In GLM and HGLM model the base premium is equal to $P_{GLM} = 18354.07$ and $P_{GLM} = 18370.64$ while in EQRM model at a much higher level $P_{EQRM} = 25167.3$. The Similar situation is in the case of structural parameters of the most rating factors which are generally higher in EQRM model while lower or similar for GLM and HGLM models. Different situation is for the factor „`RegionG`", which treated as the random effect takes much higher value compared to the value in EQRM and GLM models, see Table 1.

	EQRM	**p-value**	**GLM**	**p-value**	**HGLM**	**p-value**
`Intercept`	25	0.00	18 354.07	0.00	18 370.64	0.00
`Driver.age`	1.00	-	1.00	-	1.00	-
`Driver.age`	1.62	0.03	1.62	0.04	1.63	0.05
`Driver.age`	2.37	0.00	2.47	0.00	2.37	0.00
`Driver.age`	2.23	0.01	2.18	0.00	2.11	0.01
`Driver.age`	1.52	0.08	1.81	0.01	1.74	0.03
`Driver.age`	1.51	0.17	1.56	0.08	1.51	0.12
`Driver.age`	0.75	0.17	0.65	0.25	0.68	0.32
`RegionA`	1.00	-	1.00	-	1.09	-
`RegionB`	1.12	0.21	1.14	0.36	1.25	-
`RegionC`	0.85	0.08	0.87	0.35	0.91	-
`RegionD`	0.86	0.14	0.90	0.44	0.99	-
`RegionE`	0.67	0.26	0.61	0.27	0.90	-
`RegionF`	0.54	0.12	0.65	0.19	0.90	-
`RegionG`	0.01	0.37	0.01	0.00	0.95	-

MC.classA	1.00	-	1.00	-	1.00	-
MC.classB	1.15	0.37	1.14	0.61	1.10	0.72
MC.classC	1.67	0.09	1.66	0.02	1.44	0.11
MC.classD	1.46	0.13	1.17	0.51	1.12	0.64
MC.classE	1.66	0.04	1.26	0.30	1.17	0.51
MC.classF	2.13	0.01	1.68	0.02	1.58	0.04
MC.classG	3.72	0.19	1.94	0.25	2.11	0.24
Veh.ageA	1.00	-	1.00	-	1.00	-
Veh.ageB	0.96	0.46	0.91	0.56	0.93	0.65
Veh.ageC	0.50	0.02	0.55	0.00	0.57	0.00
Veh.ageD	0.23	0.00	0.27	0.00	0.28	0.00

Table 1 Tariff rates for EQRM, GLM and HGLM model

In order to compare models by means of a unified measure, the 5–fold cross–validation procedure was applied. RMSE error in each validation set and Cross–validation RMSE (cv) are presented in Table 2 and Table 3.

Validation set	RMSE EQRM	RMSE GLM	RMSE HGLM
ValidPart1	46440.6	44242.1	44163.6
ValidPart2	35833.4	34029.2	34005.4
ValidPart3	42178.1	34799.3	34607.6
ValidPart4	41178.4	44823.1	44870.5
ValidPart5	45142.9	48603.8	48546.5

Table 2 RMSE for EQRM, GLM and HGLM model

Model	Cross–validation RMSE
EQRM	42 154.7
GLM	41 299.5
HGLM	41 238.7

Table 3 Cross–validation RMSE for EQRM, GLM and HGLM model

For the analyzed data set the lowest error cv obtained HGLM model. Therefore in his case, for further calculations of tariff rates this model should be used. Using cross–validation procedure gives a rather demonstrative result that may be a prelude to further analysis and verification of models. The problem lies in the selection of unified tests that would allow the final choice of the method for rate–making.

6 Conclusions

In this paper we presented briefly the quantile regression and the generalized regression models. After that we analyzed the capabilities of application both models in insurance rate–making process and we computed the numerical example. In order to estimate models and than realized the cross–validation procedure, the computer implementation of some algorithms was necessary.

There are few reasons for which we tested the capabilities of the quantile regression in rate-making, a specially several important statistical advantages. Firstly a distribution of error terms is left unspecified, which is the main virtue of the method as far as robustness to outliers is concerned. Secondly quantile estimates detect the influence of covariates on alternate parts of the conditional distribution, which we can choose arbitrarily (various orders of quantile). Thus it can be recommended in cases of non–normal asymmetric distributions – asymmetric or fat–tailed distributions. Thirdly there is the possibility to take into consideration the polices with no claims by τ parameter, which is impossible for HGLM Gamma–Gamma model. That is why in our empirical example the base premium and tariff rates are higher for EQRM than in GLM and HGLM.

Acknowledgements

This research is supported by the grant of Polish Ministry of Science and Higher Education (no. NN111461540).

References

[1] De Jong, P., and Heller, G. Z.: *Generalized Linear Models for Insurance Data*. Cambridge University Press, Cambrige, 2008.

[2] Koenker, R., and Basset, B.: Regression Quantiles, *Econometrica*, **46** (1978), 33–50.

[3] Koenker, R.: *Quantile regression*. Cambridge University Press, Cambridge, 2005.
[4] Koenker, R., and Hallock, K. F.: Quantile regression. *Journal of Economic Perspectives*, **15**(4) (2001), 143–156.
[5] Kudryavtsev, A. A.: Using quantile regression for rate–making. *Insurance: Mathematics and Economics* **45** (2009), 296–304.
[6] Lee, Y., and Nelder, A. J., and Pawitan, Y.: *Generalized Linear Models with Random Effects*, Monographs on Statistics and Applied Probability 106, Chapman & Hall\CRC, 2006.
[7] McCullagh, P., and Nelder, J. A.: *Generalized Linear Models*. Chapman & Hall/CRC, New York, 1999.
[8] McCulloch, Ch. E., and Searle, Sh. R.: *Generalized, Linear, and Mixed Models*, Wiley Series in Probability and Statistics, John Wiley & Sons, INC., 2001.
[9] Ohlsson, E., and Johansson, B.: *Non–Life Insurance Pricing with Generalized Linear Models*, Springer-Verlag, Berlin, 2010.
[10] Orwat–Acedańska, A., and Trzpiot, G.: Quantile regression in management Style Analysis of mutual balanced funds (in Polish). *Financial Investments and Insurances – World Trends and Polish Market*, University of Economics in Wrocław, **183** (2011a), 415–425.
[11] Orwat-Acedańska, A., and Trzpiot, G.: The classification of Polish mutual balanced funds based on the management style – quantile regression approach. *Theory and Applications of Quantitative Methods, Econometrics*, **31**(194) (2011b), 9–23.
[12] Portnoy, S., and Koenker, R.: The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-Error Versus Absolute–Error Estimators, with Discussion, *Statistical Science*, **12** (1997).
[13] Trzpiot, G.: The Implementation of Quantile Regression Methodology in VaR Estimation (in Polish). *Studies and Researches of Faculty of Economics and Management University of Szczecin*, Szczecin, (2008), 316–323.
[14] Trzpiot, G.: Quantile Regression Model versus Factor Model Estimation. *Financial Investments and Insurances – World Trends and Polish Market* **60** (2009a), University of Economics in Wrocław, 469–479.
[15] Trzpiot, G.: Application weighted VaR in capital allocation. *Polish Journal of Environmental Studies*, Olsztyn, **18**, 5B, (2009b), 203–208.
[16] Trzpiot, G.: Estimation methods for quantile regression, *Economics Studies* **53** (2009c), Karol Adamiecki University of Economics in Katowice, 81–90.
[17] Trzpiot, G.: Quantile Regression Model of Return Rate Relation – Volatility for Some Warsaw Stock Exchange Indexes (in Polish). *Finances, Financial Markets and Insurances. Capital Market*, University of Szczecin, **28** (2010), 61–76.
[18] Trzpiot, G.: Bayesian Quantile Regression, *Economics Studies* **65** (2011), Karol Adamiecki University of Economics in Katowice, 33–44.
[19] Wolny–Dominiak, A.: Rate-making in Automobile Insurance with HGLM (in Polish), *Proceedings of I Scientific Spatial Econometrics and Regional Economic Analysis,* Lodz, 2010.
[20] Wolny-Dominiak, A., and Trzęsiok, M.: Monte Carlo Simulation Applied to A'Priori Rate Making, *Proceedings of 26th International Conference Mathematical Methods in Economics*, Liberec, 2008.

Spatial statistics in the analysis of county budget incomes in Poland with the R CRAN

Alicja Wolny-Dominiak[1], Katarzyna Zeug-Żebro[2]

Abstract. Since Waldo Tobler [13] formulated the first law of geography, which says that everything is related, but near objects are more related than distant ones, spatial modeling has become an important research area. The methods which were developed proved to be excellent tools which can also be used in regional analysis. The most common are measures of spatial autocorrelations, which show the dependence of variables in respect of spatial localization. Spatial correlation allows to determine that intensification of a given phenomenon is more perceivable in neighboring units than in units distant from each other. The main objective of this paper is to present spatial dependences analysis using measures of global and local spatial autocorrelation with a free software environment R CRAN. The analysis is carried out using the real data set of budget incomes of counties in Poland.

Keywords: spatial autocorrelation, global and local statistics, R CRAN

JEL Classification: C44
AMS Classification: 65C60

1 Introduction

Methods of spatial statistics are used to identify spatial patterns and spatial dependency. Testing occurrence of spatial dependency boils down to verify the hypothesis of the existence of spatial autocorrelation in the data spatially localized. The evaluation of spatial autocorrelation requires the knowledge of the extent and specificity of spatial diversity, i.e. diversity of characteristics of individual sites and geographic regions.

Until recently, the rare use of spatial autocorrelation measures in practice resulted from complex and time-consuming calculation procedures. For some time, however, there has been a rapid development in computer software that allows to carry out research (often very complex) in the field of spatial statistics and econometrics. One of such programs is the R CRAN, within which packages `{spdep}` [4] and `{maptools}`, used to analyze regional and spatial data dependences, are developed. R CRAN can successfully replace the familiar, expensive software because it is multifunctional and available for free.

The objective of this paper is to study spatial dependency using of global and local spatial autocorrelation measures. All calculations and maps were made in the statistical program R CRAN based on the data relating to the budget incomes of counties in Poland in 2010. The data was obtained from the Local Data Bank of the Central Statistical Office (www.stat.gov.pl).

2 Spatial statistics

There are two types of indicators of spatial associations (ISA): global and local measures of autocorrelation. The global autocorrelation follows from the existence of correlations across the spatial unit test. The local measure shows a spatial dependency the variable with neighboring units in a particular location. The most commonly used global and local measures are: the Moran statistics I [11] and the Geary statistics C [6], [1]. The spatial autocorrelation occurs when a certain phenomenon in a single spatial unit alters the probability of occurrence of this phenomenon in the neighboring units [3]. In general, the positive spatial autocorrelation occurs when we observe the accumulation, in terms of the location, high or low values of observed variables. In the case of negative autocorrelation, high values adjacent to low, and low to high, creating a kind of checkerboard [12]. The lack of spatial autocorrelation means the spatial randomness, i.e. the high and low values of observed variables are distributed independently.

[1] University of Economics in Katowice, Department of Statistical and Mathematical Methods in Economics, alicja.wolny-dominiak@ue.katowice.pl.
[2] University of Economics in Katowice, Department of Mathematics, katarzyna.zeug-zebro@ue.katowice.pl.

2.1 Selected global statistics

The Moran statistics is one of the most widely used measures in the study of spatial autocorrelation. The Global Moran's I is defined as follows:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \cdot \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} = \frac{n}{S_0} \cdot \frac{z^{\mathrm{T}} W z}{z^{\mathrm{T}} z} \tag{1}$$

where: x_i, x_j are the values of variables in spatial unit *i* and *j*, $\bar{x}$ is the mean of variable for all units, *n* is the total number of spatial units that are included in the study, S_0 is the sum of all elements of a spatial weight matrix, *z* is a column vector of elements $z_i = x_i - \bar{x}$, *W* is the spatial weight matrix degree *n*, defining the structure of the neighborhood, w_{ij} is an element of weights matrix *W* [10]. This statistic takes values ranging from $[-1,1]$: positive, when tested objects are similar, negative, when there is no similarity between them and approximately equal to 0 for a random distribution of objects.

Cliff and Ord [5] have shown that the distribution of Moran statistics is asymptotically normal. Thus, the statistical significance of spatial autocorrelation can be verified using normalized statistics : $I_S \sim N(0,1)$:

$$I^S = \frac{I - \mathrm{E}(I)}{\sqrt{\mathrm{Var}(I)}} \tag{2}$$

where: $\mathrm{E}(I)$ is the expected value of Moran's and $\mathrm{Var}(I)$ is its variance:

$$\mathrm{E}(I) = -\frac{1}{n-1},\ \mathrm{Var}(I) = \frac{n^2 S_1 - n S_2 + 3S_0^2}{\left(n^2-1\right)S_0^2} - \frac{1}{(n-1)^2} \tag{3}$$

$$S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\ ,\ S_1 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(w_{ij} + w_{ji}\right)^2\ ,\ S_2 = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} w_{ij} + \sum_{j=1}^{n} w_{ji}\right)^2 \tag{4}$$

If the Moran statistic has a value $I \approx -(n-1)^{-1}$, $I^S \approx 0$ it indicates a random spatial pattern. However, when $I > -(n-1)^{-1}$, $I^S > 0$ the spatial autocorrelations is positive, and if $I < -(n-1)^{-1}$, $I^S < 0$, the spatial autocorrelations is negative.

Another global measure of spatial autocorrelation is Global Geary's *C*. This statistic, is given by

$$C = \frac{(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(x_i - x_j\right)^2}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \cdot \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} = \frac{n}{(n-1)}\left[\frac{n}{S_0} \cdot \frac{z^{\mathrm{T}}\mathrm{diag}(w_{i.})z}{z^{\mathrm{T}} z} - I\right] \tag{5}$$

where all elements of the formula are defined as in statistic *I*. The above formula shows that the Geary measure can be expressed by the Moran statistic [8]. Although Moran and Geary measures give similar results, the Moran statistic is more effective. This is due to greater sensitivity of the variance of the Geary statistic to the distribution of sample. Values of this statistic can be impaired when the matrix of weights is asymmetrical. In order to verify the hypothesis of no spatial correlation, the Geary statistic can be standardized:

$$C^S = \frac{C - \mathrm{E}(C)}{\sqrt{\mathrm{Var}(C)}} \sim N(0,1) \tag{6}$$

where: $\mathrm{E}(C)$ is the expected value of Geary's and $\mathrm{Var}(C)$ is its variance:

$$\mathrm{E}(C) = 1,\ \mathrm{Var}(C) = \frac{(n-1)(2S_1 + S_2) - 4S_0^2}{2(n+1)S_0^2} \tag{7}$$

The value of Global Geary's *C* is always positive and takes values ranging from $[0, 2]$. In the case, of: $1 < C < 2$, $C^S > 0$, the spatial autocorrelation is negative; when $0 < C < 1$, $C^S < 0$, the spatial autocorrelation is positive; finally, when $C \approx 1$, $C^S \approx 0$, there is no spatial autocorrelation.

2.2 Selected local statistics

We can use local indicators of spatial association (LISA), a Local Moran statistics and a Local Geary statistics, to identify spatial systems. The Local Moran determines clusters of spatial units and studies whether the unit is surrounded by neighboring units with similar or different values of the variable studied in relation to the random distribution of these values in the studied space [10].

In the case of non-standardized values of the variable and row-standardized spatial weight matrix [2] ($\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} = n$), the local Moran is given by:

$$I_i = \left[(x_i - \bar{x}) \sum_{j=1}^{n} w_{ij} (x_j - \bar{x}) \right] \Bigg/ \left[\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n} \right] \tag{8}$$

where all elements of the formula are defined as in the Global Moran's *I*. The standardized Local Moran's I_i^S is used to test the statistical significance of local spatial autocorrelation [1]:

$$I_i^S = \frac{I_i - \mathrm{E}(I_i)}{\sqrt{\mathrm{Var}(I_i)}} \sim N(0,1) \tag{9}$$

where: $\mathrm{E}(I_i)$ is the expected value of the Local Moran and $\mathrm{Var}(I_i)$ is its variance

$$\mathrm{E}(I_i) = -\frac{\sum_{j=1}^{n} w_{ij}}{n-1} \text{ i } \mathrm{Var}(I_i) = \frac{(n-k)\sum_{i\neq j} w_{ij}^2}{n-1} + \frac{2(2k-n)\sum_{l\neq i}\sum_{hi\neq i} w_{il} w_{ih}}{(n-1)(n-2)} - \left(\frac{-\sum_{i\neq j} w_{ij}}{n-1} \right)^2 \tag{10}$$

where $k = \left(\frac{1}{n}\sum_i (x_i - \bar{x})^4 \right) \Big/ \left(\frac{1}{n}\sum_i (x_i - \bar{x})^2 \right)^2$.

When I_i^S is negative, the spatial autocorrelation is negative too, i.e. when the object is surrounded by spatial units with significantly different values of the studied variable. The spatial autocorrelation is positive when $I_i^S > 0$, the object is surrounded by similar neighboring units.

According to Anselin [1] a Local Geary statistics for an observation *i* may be defined as

$$C_i = \sum_{j\neq i}^{n} w_{ij} (z_i - z_j)^2 \tag{11}$$

where $z_i = x_i - \bar{x}$, $z_j = x_j - \bar{x}$ and w_{ij} are the elements of the row-standardized binary symmetric spatial weight matrix **W**. The test statistic for C_i^S is

$$C_i^S = \frac{C_i - \mathrm{E}(C_i)}{\sqrt{\mathrm{Var}(C_i)}} \sim N(0,1) \tag{12}$$

where: $\mathrm{E}(I_i)$ is the expected value of the Local Moran and $\mathrm{Var}(I_i)$ is its variance

$$\mathrm{E}(C_i) = \frac{n\sum_{j=1}^{n} w_{ij} \cdot \sum_{j=1}^{n} (z_i - z_j)^2}{(n-1)^2} \text{ i } \mathrm{Var}(C_i) = \frac{\left[(n-1)\sum_{i=1}^{n} w_{ij}^2 - \left(\sum_{j=1}^{n} w_{ij} \right)^2 \right] \cdot \left[(n-1)\sum_{j=1}^{n} (z_i - z_j)^4 - \left[\sum_{j=1}^{n} (z_i - z_j)^2 \right]^2 \right]}{(n-1)^2 (n-2)} \tag{13}$$

the significant testing on local spatial association can be conducted based on the calculated test statistics above. The C_i statistic is interpreted in the same way as the Local Moran.

3 ISA for incomes of counties in Poland with R

In the empirical example, we analyzed global and local indicators of spatial associations (ISA) for incomes of counties in Poland in 2010 year. For all computations and maps we used the free software environment R. We started with the calculation of the spatial weight matrix for 376 counties in Poland which measures spatial links between objects. This matrix is necessary to analyze the neighborhood. Based on the weight matrix we computed the neighborhood matrix according to adjacency criteria. The neighborhood map of counties is as follows:

number of links	number of counties	% of counties
1	31	8.24%
2	12	3.19%
3	21	5.59%
4	41	10.90%
5	90	23.94%
6	77	20.48%
7	57	15.16%
8	29	7.71%
9	11	2.93%
10	5	1.33%
11	2	0.53%

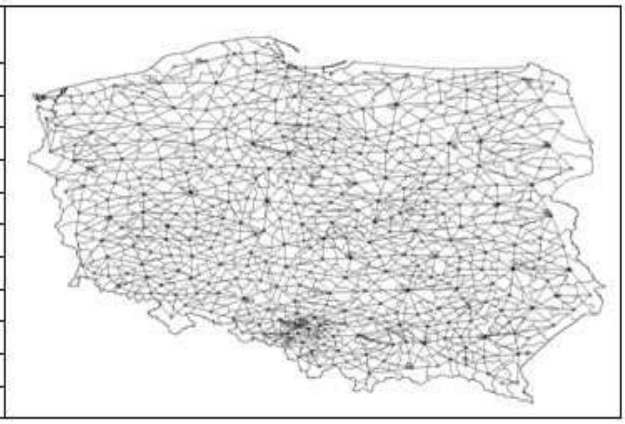


Figure 1 Distribution of the number of links for the counties and the neighborhood matrix for Poland

The number of nonzero links in all counties is equal to 1996 and the average number of links for one county is 5.31. As we can see there are 31 least connected counties with 1 link and 2 most connected counties with 11 links. Over 23% of counties have five neighbors.

R code:
```
> map<-readShapePoly("C:/DANE/POL/POL_adm2.shp")
> map.nb<- poly2nb(as(map,"SpatialPolygons"))
> map.listw<-nb2listw(map.nb, style="W")
> coord=coordinates(map)
> plot(map.nb,coord, add=TRUE)
```

Then we calculated Moran's I global statistics using the test under randomization and $I = 0.0615$ with expectation $E(I) = -0.0027$, variance $Var(I) = 0.0005$. The small p-value at 0.0025 shows significance of the statistics. The value of Moran I is close to zero, which indicates no spatial autocorrelation. This means that there is no similarity between neighboring counties in terms of incomes.

R code:
```
> moran<-moran.test(data$Income, map.listw)
> moran.plot((data$Income-mean(data$Income))/sd(data$Income),map.listw, xlab="Income budget of counties in Poland ", ylab="Spatial lags for Income")
```

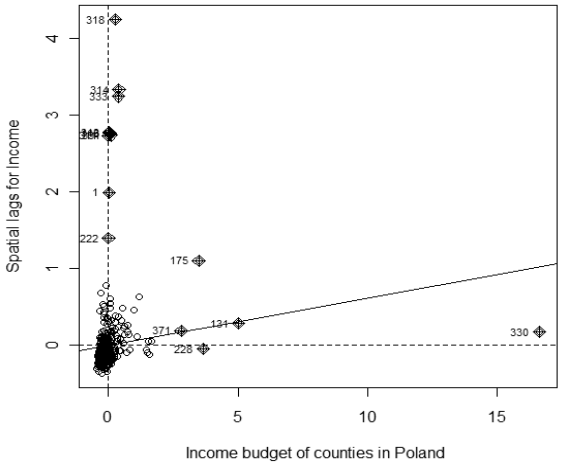


Figure 2 Scatter plot for the Moran global statistic

We also computed global Geary's statistics using the test under randomization. The results are similar to Global Moran statistics except for a p-value, which is equal to 0.5062. The value $C = 1.0031$ is near one which indicates no spatial autocorrelation, but p-value proves that Global Geary's statistics is insignificant.

In order to analyze the spatial autocorrelation in every county we calculated Local Moran statistics and Local Geary statistics. First of all, we tested significance of both statistics. The results are shown on the following maps:

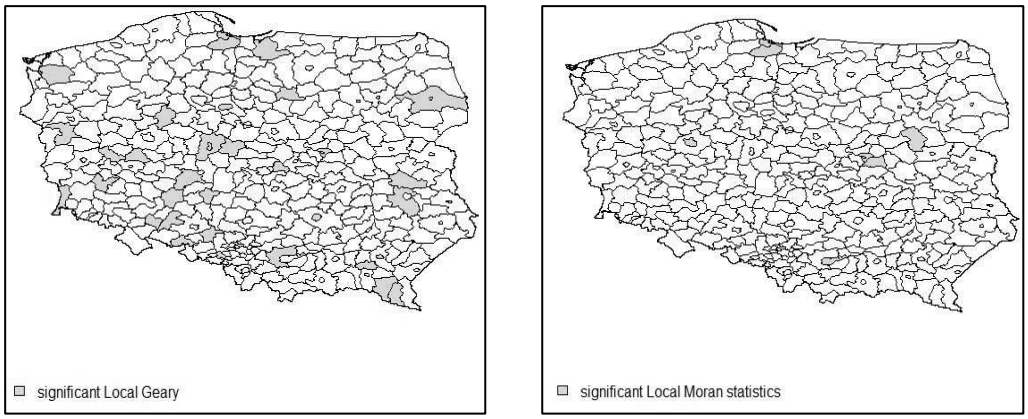


Figure 3 Counties with significant Local Geary and Local Moran statistics

The Local Moran is significant only for 8 counties: “Powiat m. Krakow”, “Powiat m. Poznań”, Powiat piaseczynski”, “Powiat pruszkowski”, “Powiat m. Warszawa”, „Powiat wołomiński”, „Powiat m. Gdańsk”, „Powiat m. Gdynia”. For all these counties Local Moran is significantly positive with the p-value below 0.05 which means that those counties are surrounded by objects with similar value of incomes, but we can-not say which counties are rich or poor. The Local Geary is significant for a greater number of counties against Local Moran. The interpretation is similar.

R code:

```
> moran.local<-localmoran(data$Income, map.listw)
> sig<-ifelse(moran.local[,5]<=0.05,"*"," insig ")
> break=c(0.0000000000000001, 0.05, 0.95, 0.999999999999999999)
> colors=cm.colors(1:3, alpha=1)
> moran.local.df=as.data.frame(moran.local)
> plot(map,col=colors[findInterval(moran.local.df[,5],break)])
> legend("bottomleft", legend=c("significant Local Moran"), fill=colors, bty="n")
> Gi.local<-localG(data$Income, map.listw)
> Gi.local.df<-as.data.frame(as.vector(Gi.local))
> sig<-ifelse(as.data.frame(as.vector(Gi.local))<=-3.083|as.data.frame(as.vector(Gi.local))<=3.083,"*"," insig ")
> colors=cm.colors(1:3, alpha=1)
> plot(map,col=colors[findInterval(Gi.local.df[,5],break)])
> legend("bottomleft", legend=c("significant Local Geary"), fill=colors, bty="n")
```

Based on Local Moran, we identified spatial regimes which show counties and neighbors with high and low values of incomes.

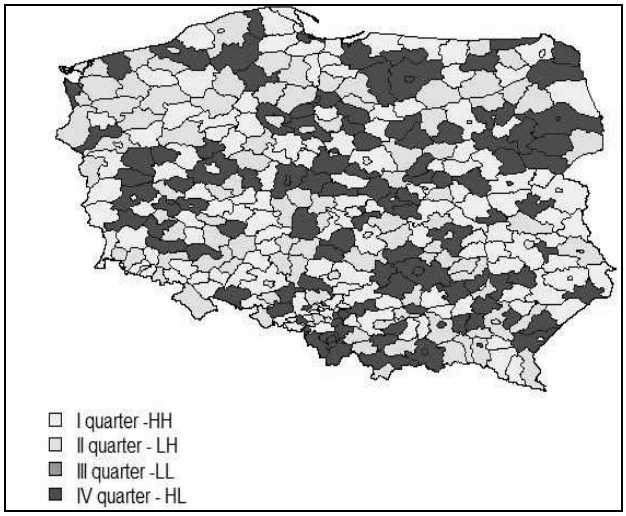


Figure 4 Spatial regimes

The classification of spatial regimes is as follows:

(i) I quarter HH – Cluster " I have high budget incomes and so do my neighbors"
(ii) II quarter LH – Outlier " I have low budget incomes among neighbors with high incomes"
(iii) III quarter LL – Cluster " I have low budget incomes and so do my neighbors"
(iv) IV quarter HL – Outlier " I have high budget incomes among neighbors with low incomes"

R code

```
> Z<-(dane$X15-mean(dane$X15))/sd(dane$X15)
> lag.Z<-lag.listw(map.listw, Z)
> q1<-ifelse(Z>0&lag.Z>0,1,0)
> q2<-ifelse(Z>0&lag.Z<0,2,0)
> q3<-ifelse(Z<0&lag.Z>0,3,0)
> q4<-ifelse(Z<0&lag.Z>0,4,0)
> q<-q1+q2+q3+q4
> q.data<-as.data.frame(q)
> break=c(1,2,3,4)
> colors=rev(heat.colors(4))
> plot(map,col=colors[findInterval(q.data$q,break)], forcefill=FALSE)
> legend("bottomleft", legend=c("I quarter -HH", "II quarter - LH", "III quarter -LL", "IV quarter - HL"),
fill=colors, bty="n")
```

4 Conclusions

Spatial methods are used increasingly frequently in the analysis of economic processes. One of the reason is the fact that spatial autocorrelation local and global measures, informing about the type and strength of spatial dependency, allow on: fuller use of the measure; to determine the relationship between reference entities; to define spatial structures [9]. Additionally, there are rapid developments in software that offers computational procedures in the field of spatial statistics and econometrics. Their effects can be observed, inter alia, in the R CRAN, which is useful for all professionals and scientists dealing with the analysis of spatial data.

References

[1] Anselin, L.: *Local Indicators of Spatial Association*-LISA. Geographical Analysis, 27, (1995), 93-115.
[2] Arbia, G.: *Spatial Econometrics: Statistical Foundations and Applications to Regional Growth Convergence*. Springer, New York, (2006).
[3] Bivand, R.: Autokorelacja przestrzenna a metody analizy statystycznej w geografii. In: *Analiza regresji geografii* (Chojnicki, Z., ed), PWN, Poznań, (1980), 23-38.
[4] Bivand, R.: *Spatial Econometrics Functions in R: Classes and Methods*. Journal of Geographical System, (2003)
[5] Cliff, A.D., Ord, J.K.: *Spatial Autocorrelation*. Pion, London, (1973).
[6] Geary, R.: *The Contiguity Ratio and Statistical Mapping*. The Incorporated Statistician, 5, (1954), 115–145.
[7] Getis, A., Ord, J.K.: *The Analysis of Spatial Association by Use of Distance Statistics*. Geographical Analysis, 24, (1992), 189–206.
[8] Griffith, D.A.: *Spatial Autocorrelations and Spatial Filtering*. Springer, Berlin-Heidelberg, (2003).
[9] Janc, K.: Zjawisko autokorelacji przestrzennej na przykładzie statystyki I Morana oraz lokalnych wskaźników zależności przestrzennej (LISA) – wybrane zagadnienia metodyczne. In: *Idee i praktyczny uniwersalizm geografii. Dokumentacja Geograficzna* (Komornicki, T., Podgórski, T., eds), 33, (2006) 76–83.
[10] Kopczewska, K.: Ekonometria *i statystyka przestrzenna z wykorzystaniem programu* R CRAN. Cedewu.pl, Warszawa, (2006).
[11] Moran, P. A. P.: *Notes on Continuous Stochastic Phenomena*. Biometrika 37 (1), (1950), 17–23.
[12] Suchecki, B. (red.): *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*. Wydawnictwo C.H. Beck, Warszawa, (2010).
[13] Tobler, W.: *A computer model simulating urban growth in Detroit region*. Economic Geography, 46(2), (1970), 234–240.

Synchronization of business cycles between Poland, the euro zone and the new member states of the European Union

Joanna Wyrobek[1], Zbigniew Stanczyk[2], Marek Zachara[3]

Abstract. In the following paper, the business cycles between Poland and other members of the European Union are assessed for the purpose of determining their dependence. In order to identify the said dependence, various methods have been employed, primarily a spectral analysis and a co-integration. A special focus is applied to two groups of the member states: the euro zone, due to its significance in the European economy, but also other states that joined the EU with Poland during the same period. Interestingly, the coherence between the Polish and the euro zone business cycles is high, while it is much lower between Poland and the other new member states. The data also proves that Poland had benefited from the accession to the EU, even though this happened during an unfavorable period of the business cycle of the euro zone's countries. The analysis performed illustrates that the Polish business cycle over the years is becoming closer to the euro zone's cycle, while remaining behind it in phase. This conclusion is vital to some companies, as it might help them in mid-term strategy planning.

Keywords: business cycle, euro zone, spectral analysis.

JEL classification: C32, F41
AMS classification: 62M15, 93E11, 91B84

1 Introduction

The purpose of this paper was to analyse the business cycle in Poland and compare it to business cycles of: the euro area (as a whole), the European Union (as a whole), and of selected EU countries, in an attempt to verify if there is an economic evidence of synchronization of the Polish economy with the euro area business cycles which would vindicate Poland's decision to join the euro zone in the nearest future. Business cycle analysis is performed using methods applicable for time-series [5], specifically spectral and co-integration methods are utilised to analyse the business cycles and the relations between them. The results from previous research in this area are briefly described in the second part, and the final part contains the calculations and results of the research conducted by the authors of the paper.

2 Methods used in the article

2.1 Christiano-Fitzgerald band-pass filter

The Christiano-Fitzgerald band-pass filter is used to extract the cyclical part of the time series. The filter was chosen because of its applicability to almost all time series and its advantages over other methods (it takes into account stochastic structure of the decomposed variable, removes all non-seasonal fluctuations, etc.). The Christiano-Fitzgerald filter requires testing of the stationarity of time series (time series can be: stationary I(0), trend stationary or non-stationary I(1)). The filter requires the removal of time-trend (if it is present) and for processes stationary at I(1) one must remove a drift (if it is present) [9].

[1]University of Economics, Cracow, Poland, wyrobekj@uek.krakow.pl
[2]University of Economics, Cracow, Poland, stanczyz@uek.krakow.pl
[3]AGH University of Science and Technology, Cracow, Poland, mzachara@agh.edu.pl

The idea of calculating the cyclic component $\hat{y}_t^c$ in the band pass filter is based on the following formula from [8]:

$$\hat{y}_t^c = \hat{B}_t(L)y_t, \quad where: \hat{B}_t(L) = \sum_{j=-(T-t)}^{t-1} \hat{B}_{j,t}L^j \quad for \quad T = 1, 2, ...T \tag{1}$$

A set of weights $\hat{B}_{j,t}$ is the solution of the equation:

$$\underset{\hat{B}_{j,t} = -(T-t), ..., t-1}{min} \int_{-\pi}^{\pi} \left| B\left(e^{-i\omega}\right) - \hat{B}_t\left(e^{-i\omega}\right) \right|^2 S_y(\omega)d\omega \quad for \quad t = 1, 2, ...T \tag{2}$$

For the CF filter for the I(1) series there is an additional (limiting) condition that: $\sum_{j=-(T-t)}^{t-1} \hat{B}_{j,t} = 0$ for $t = 1, 2, ..., T$ which provides for removal by filter of a stochastic trend. Further detail on the filter can be found in [8] and [9].

2.2 Single spectrum analysis method

The origin of spectral analysis is based on the idea of representing time series as the sum of sinusoids at various frequencies (cycles). Spectral analysis of cyclic data requires the Fourier transform, which is used to transform the time domain representation of the series into the frequency domain representation of the series. In order to determine the significance of different frequencies in data one calculates a spectrogram. A spectrogram displays the power of a signal as a function of both: time and frequency simultaneously.

According to [9], power spectrum of a stochastic process with discrete time $\{y_t\}_{t=-\infty}^{+\infty}$ with a zero mean and stationary covariance function is defined as the Fourier transform of autocovariance series $\{y_t^v\}_{k=-\infty}^{+\infty}$ of this process and is given as:

$$S_y(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{+\infty} \gamma_k^y \cdot e^{-i\varpi k} \quad for \quad \omega \in (-\pi, \pi) \tag{3}$$

Due to the fact that the spectrogram calculated by using the above method is very "fuzzy", certain methods are used to reduce this variability (smoothing methods), and one of the most popular is the Parzen window. The power spectrum estimator then takes the form:

$$\hat{S}(\omega) = \frac{1}{2\pi} \sum_{k=-H}^{H} w_k \cdot \hat{\gamma}_k^y \cdot e^{-i\varpi k} = \frac{1}{2\pi} \left[W_o \hat{\gamma}_0^y + 2 \sum_{k=1}^{H} w_k \cdot \hat{\gamma}_k^y \cdot cos(\omega k) \right] \tag{4}$$

Where empirical autocovariances are

$$\hat{\gamma}_k^y = \frac{1}{T} \sum_{t=1+k}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y}) \quad for \quad k = 0, 1, .., T-1 \tag{5}$$

and Parzen window weights are:

$$w_k = \begin{cases} 1 - 6(k/H)^2 + 6(|k|/H)^3 & for \quad |k| \leq H/2 \\ 2(1 - |k|/H)^3 & for \quad H/2 \leq |k| \leq H \\ 0 & for \quad |k| > H \end{cases} \tag{6}$$

Maximum allowable lag time for Parzen window, called the truncation lag is chosen according to the rule $H = int(2\sqrt{T})$.

2.3 Cross-spectral analysis

"Cross spectral analysis allows one to determine the relationship between two time series as a function of frequency. Normally, one supposes that statistically significant peaks at the same frequency have been shown in two time series as that we wish to see if these periodicities are related with each other and, if so, what the phase relationship (time shift) is between them [6]".

There are several methods of calculating the cross-spectrum, one of which is given by Bloomfield [2]. The time series X and Y can first be "combined" in the time domain (before the Fourier transform) by calculating the lagged cross-covariance function. The resulting function is then subjected to a Fourier transform and a cross spectrum periodogram is obtained. Cross-covariance can be written as: $c_{x,y,r} = \frac{1}{n}\sum x_t y_{t-r}$, where t and $t-r = 0,\ 1,\ 2,\ ..,\ n-1$, and r is the time lag of one series relative to the other. The Fourier transform is then carried out to obtain the cross-spectrum periodogram:

$$I_{x,y}(\omega) = \frac{1}{2\pi}\sum_{|r|<n}(c_{x,y,r}\cdot e^{-ir\omega}) \tag{7}$$

Similarly to the single spectrum periodogram (spectrogram), the cross-spectrum periodogram is also smoothed, e.g. by a Parzen window. Applying the cross-spectrum analysis it is possible to calculate the following three measures: squared coherence, gain value and time shift between the series.

According to [9] a stochastic process with discrete time $\{x_t\}_{t=-\infty}^{+\infty}$, with zero mean and stationary covariance function is an independent variable, whereas the similar process $\{y_t\}_{t=-\infty}^{+\infty}$ is the dependent variable, the cross power spectrum (cross-spectral density) of these variables is defined as a Fourier transform of the cross-covariance series of these variables and is given by the formula:

$$S_{yx}(\omega) = \frac{1}{2\pi}\sum_{k=-\infty}^{+\infty}\gamma_k^{yx}e^{-i\omega k} = c_{yx}(\omega) - iq_{yx}(\omega) \quad for \quad \omega\in[-\pi,\pi] \tag{8}$$

where $c_{yx}(\omega) = \frac{1}{2\pi}\sum_{k=-\infty}^{+\infty}\gamma_k^{yx}cos(\omega k)$ is called co-spectrum and is a real part of cross-spectrum, while $q_{yx}(\omega) = \frac{1}{2\pi}\sum_{k=-\infty}^{+\infty}\gamma_k^{yx}sin(\omega k)$, called the quadrature spectrum, is a negative imaginary part of the cross-spectrum. It is possible to calculate three cross-spectral statistics on the basis of cross power spectrum: gain value ($G_{yx}(\omega)$), phase shift ($\Phi_{yx}(\omega)$), and squared coherence ($K_{yx}^2(\omega)$):

$$\begin{aligned} G_{yx}(\omega) &= \frac{\sqrt{c_{yx}^2(\omega)+q_{yx}^2(\omega)}}{S_x(\omega)} \\ \Phi_{yx}(\omega) &= tan^{-1}\left(\frac{-q_{yx}(\omega)}{c_{yx}(\omega)}\right) \\ K_{yx}^2(\omega) &= \frac{c_y^2x(\omega)+q_y^2x(\omega)}{S_y(\omega)S_x(\omega)} \end{aligned} \tag{9}$$

where $S_x(\omega)$ is the power spectrum of the process $\{x_t\}$, while $S_y(\omega)$ is the power spectrum of the process $\{y_t\}$.

2.4 Correction for the Polish time series used in the article

Countries and regions examined in this analysis (data was already seasonally adjusted with Demetra RSA1 [4] method) were subject to the transformations described above; i.e. the time series had time trend removed (or drift removed) if needed. Such data were then used for the Christiano-Fitzgerald filter, and finally for spectral analysis. However, for Poland, an additional adjustment was necessary; i.e. the elimination of an additional business cycle peak resulting from Polish accession to the European Union in 2004 (Fig. 1).

The implementation of the correction for the Polish series was preceded by analysing the results of calculations for an uncorrected series. Adjustments do not distort the picture of cyclical changes in Poland, because except for EU accession, Poland has had the same cyclical troughs and peaks as other EU countries, and therefore the correction erased the only anomaly that could disrupt the spectral analysis.

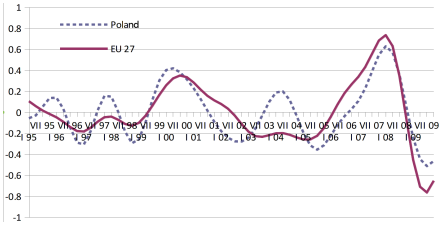


Figure 1 Polish and EU cyclical parts of GDP (percentage changes in real GDP with seasonality and time trend removed, and subjected to the CF filter). Calculations based on the World Bank real GDP data [11].

3 Brief overview of previous studies

The subject discussed in this paper has already received some attention, however the conclusions obtained by researchers differ substantially from each other. This is partly the result of different research methods, and partly due to the use of different underlying data provided by different sources. Several prominent studies and their results has been presented in Table 1.

Authors	method	Poland		Euro area		Germany	
		short cycle	long cycle	Coherence	Phase Shift	Coherence	Phase Shift
Adamowicz [1]	Baxter-King			0.39	-1.26	0.43	-1.17
Adamowicz	Christiano-Fitzgerald			0.47	-0.99	0.46	-0.99
Adamowicz	Hodrick-Prescott			0.35	-1.18	0.41	-0.96
Adamowicz	structural model			0.06	-3.96	0.08	-4.85
Skrzypczyski [9]	Christiano-Fitzgerald	3 y.	6-7 y.	0.52	-1 to -4		
Skrzypczyski [10]	filter CF		6-7 y.				
Skrzypczyski [10]	SVAR		12 y.				

Table 1 Research results on the length of the business cycles in Poland and their relationship to cycles Germany and the euro area (measured on the basis of the cyclical part of GDP)

Other research worth mentioning is [3] who used wavelets for calculating business sycle synchronization and concludes that the Polish business cycle is becoming more and more synchronized with the cycle of the euro area. On the other hand in [7] finds on average a lower degree of synchronization of the euro area cycle with the cycles of the countries of Central and Eastern Europe compared with the cycles of the member countries of the euro area.

4 Strength of the relationship between business cycles in Poland and cycles in other EU countries

The strength of the relationship between cycles (in addition to the length of the business cycle) of a particular country with other countries may indicate a strong relationship between their economies. In the case of spectral analysis, the strength of the relationship between cycles is measured by the squared coherence; the higher the coherence, the stronger the relationship.

As can be seen in Table 2, Polish economy cycles show a strong similarity to cycles in countries such as Finland, the Netherlands, Sweden, the United Kingdom (and also the whole EU). For most of the examined frequencies, coherence ratios were often over 80%. It can be also observed that for all countries there is high coherence in respect of long cycles (longer than 15 quarters).

A surprisingly low coherence exists between Poland and the majority of its neighbors; e.g. the Czech Republic, Slovakia or the Baltic States. It is also interesting to note that in spite of a very strong import-export ties between Poland and Germany, coherence between the two countries (which is quite high) is a little lower than the coherence of the Polish economy with the whole of the European Union and the euro area. These differences are even more apparent when the medians are considered. For the period

1995-2009, the median coherence with Germany is 72.3%, while it is 80% with the euro area and 82.1% with the European Union. The differences, which invite some more research, can be attributed to the effect of Polands accession to the EU in 2004, and to the exceptionally good performance of the Polish economy in 2009.

Country	Average	30	20	15	12	10	8,57	7,5	6,676	6
Austria	51.16	84.80	76.10	72.20	46.70	46.40	60.50	39.00	14.40	20.30
Belgium	49.93	65.10	51.30	57.70	35.40	41.70	63.40	59.90	49.90	25.00
Czech	46.1	63.90	55.20	58.80	50.70	42.40	14.00	1.20	44.70	84.00
Estonia	52.22	63.40	43.70	36.90	64.20	84.70	71.50	14.00	26.20	65.40
Finland	61.61	87.40	82.90	82.00	82.70	83.90	68.20	17.30	6.00	44.10
France	64.97	85.20	80.80	82.30	85.40	87.30	80.10	45.60	9.80	28.20
Germany	64.88	79.60	69.20	72.30	85.00	89.00	74.40	27.30	21.20	65.90
Hungary	63.46	88.40	83.80	73.40	72.70	80.60	65.20	22.50	16.70	67.80
Ireland	58.8	80.00	71.80	73.60	85.80	89.40	77.20	33.80	15.60	2.00
Italy	58.93	76.90	67.50	68.50	81.10	88.50	77.70	33.40	8.60	28.20
Latvia	51.26	77.40	60.00	52.60	62.00	76.70	71.80	33.40	8.40	19.00
Lithuania	45.1	49.90	30.90	33.70	58.60	73.30	59.20	17.80	21.00	61.50
Netherlands	66.67	91.70	88.40	84.60	80.40	77.20	71.00	38.80	8.80	59.10
Portugal	56.98	81.30	72.50	67.80	45.80	43.60	70.60	71.20	30.90	29.10
Slovakia	27.31	37.00	30.10	37.90	45.90	35.10	10.90	6.20	17.60	25.10
Slovenia	55.91	85.20	76.70	72.50	64.90	66.30	63.90	22.20	7.50	44.00
Spain	61.9	90.10	84.90	82.30	72.90	71.60	73.70	50.20	2.80	28.60
Sweden	63.39	90.00	90.40	83.60	77.10	86.50	76.40	25.20	6.00	35.30
UK	65.94	86.30	84.40	84.90	81.30	79.90	70.70	30.20	12.10	63.70
EU 27	67.28	86.50	83.00	82.10	83.80	86.80	76.30	30.00	13.80	63.20
Euro 17	67.66	86.50	81.50	80.00	83.70	87.90	78.40	33.00	14.00	63.90

Table 2 Squared coherence between business cycle in Poland and other European countries, [in %] (for different cycle lengths); years 1995-2009. Based on the World Bank real GDP data [11].

5 Phase shift of the Polish business cycles in relation to business cycles in other EU countries

Table 3 presents the calculated phase shifts (difference in time) between business cycles of the Polish economy and cycles of other countries (for different frequencies). The negative sign means that Poland's cycle is ahead of a particular country. It is important to note that the shifts are expressed in quarters of a year, therefore a shift of 20 means 5 year time difference. As can be seen in this table, Polish cycle is generally behind most of the EU countries, following their cycle pattern. This has changed in recent years, since until 2005 Poland was often ahead of other EU countries. This may be regarded as a proof that Polish economy is strengthening its ties with the rest of the Europe and is more and more influenced by the state of the joined economies.

6 Conclusions

When comparing Poland to the euro area countries, a short cycle for Poland (estimated at 3 - 4 years, or about 10 quarters) is similar to the short cycle of the strongest euro economies (Germany, France, Italy). However, the long cycle for Poland (estimated at 6 - 7 years, approximately 20 - 30 quarters) tends to be a little longer than the long cycle of these strongest euro economies

For the period 1995 - 2009 coherences are significantly high for virtually all countries, and high coherence is observed even for short cycles (even for 6 quarter cycles). A surprisingly low coherence exists between Poland and the majority of its neighbors; i.e. the Czech Republic and Slovakia and the Baltic States.

The amplitudes of fluctuations of the Polish economy are bigger than the amplitudes of fluctuations in the euro zone and in the major economies in the euro zone, but comparing the periods 1991-2005 and 1995-2006 the similarity of the amplitudes have increased. The same applies to most new EU members with the exception of the Czech Republic which had relatively low amplitudes of business cycles.

All the results indicate that the synchronization between the business cycle in Poland and in the euro area is relatively high and appears to be increasing. However, it should be added that the results could be to some extent influenced by the conduct of economies during the last global crisis of 2008.

Country	30	20	15	12	10	8.6	7.5	6.7	6
Austria	-0.06	0.81	1.07	0.43	0.25	0.63	0.89	1.03	0.03
Czech	-1.38	-0.48	-0.17	-0.54	-0.79	-0.04	0.27	-1.23	-0.97
Estonia	1.55	0.75	0.75	0.55	0.28	0.39	0.55	1.62	-1.31
Finland	0.93	0.29	-0.02	-0.09	0.09	0.40	0.65	-1.25	-1.13
France	0.29	0.66	0.73	0.47	0.53	0.73	0.93	0.89	-0.29
Germany	0.01	0.24	0.49	0.32	0.21	0.38	0.44	-0.89	-1.00
Hungary	1.43	0.89	0.59	0.02	-0.04	0.32	0.48	-0.95	-1.07
Ireland	0.78	0.58	0.62	0.51	0.49	0.66	1.11	0.95	-0.49
Italy	0.29	0.39	0.48	0.31	0.17	0.24	0.48	0.72	-0.07
Latvia	0.14	0.24	0.60	0.24	0.01	0.32	0.74	0.32	1.22
Lithuania	-0.25	-0.72	-0.46	-0.46	-0.36	0.16	0.58	-1.50	-1.23
Netherlands	-0.31	-0.09	-0.08	-0.21	0.06	0.51	0.74	-0.06	-0.79
Portugal	-0.50	0.38	0.63	0.46	1.08	1.04	0.83	0.76	0.21
Slovakia	-2.55	-1.59	-0.76	-0.73	-0.46	0.75	1.85	0.91	1.48
Slowenia	-0.74	-0.43	-0.16	-0.29	-0.03	0.47	0.69	-1.06	-1.06
Spain	-0.97	-0.12	0.38	0.32	0.69	0.97	1.03	0.72	-0.94
Sweden	2.46	1.58	1.13	0.47	0.18	0.33	0.47	-0.96	-0.88
UK	1.29	0.65	0.38	0.05	-0.02	0.36	0.63	-1.17	-1.21
EU 27	0.56	0.45	0.43	0.19	0.16	0.38	0.52	-0.73	-0.94
EURO 17	0.35	0.37	0.40	0.20	0.19	0.38	0.50	-0.55	-0.85

Table 3 Phase shifts (in quarters) between business cycle in Poland and other countries (for different frequencies); years 1995-2009. Calculations based on the World Bank real GDP data [11].

References

[1] Adamowicz, E [et al.]: Synchronizacja cyklu koniunkturalnego polskiej gospodarki z krajami strefy euro w kontekscie struktury tych gospodarek, In: *Raport na temat pelnego uczestnictwa Rzeczypospolitej Polskiej w trzecim etapie Unii Gospodarczej i Walutowej*, NBP, Warszawa, 2009.

[2] Bloomfield, P.: Fourier analysis of time series, Wiley, New York, 1976.

[3] Bruzda, J.: Business cycle synchronization according to wavelets the case of Poland and the euro zone member countries, *Bank i Kredyt*, **42 (3)**, 2011, 5–32

[4] Gomez, V., Maravall, A.:, Programs TRAMO and SEATS - Instructions for the User, Italian National Institute of Statistics, *http://www3.istat.it/strumenti/metodi/destag/software/manualdos.pdf*, 1997.

[5] Hamilton, J.D.: Time series analysis, Princeton University Press, Princeton, 1994.

[6] Hartmann, D.: Objective Analysis. University of Washington, *http://www.atmos.washington.edu*

[7] Konopczak, K.: Analiza zbieznosci cyklu koniunkturalnego gospodarki polskiej ze strefa euro na tle krajow Europy Srodkowo-Wschodniej oraz panstw czlonkowskich strefy euro, In: *Raport na temat pelnego uczestnictwa Rzeczypospolitej Polskiej w trzecim etapie Unii Gospodarczej i Walutowej*, NBP, Warszawa, 2009.

[8] Nilsson, R., Gyomai G.: Cycle extraction, OECD, *http://www.oecd.org/dataoecd/32/13/41520591.pdf*, 2011.

[9] Skrzypczynski, P.: Wahania aktywnosci gospodarczej w Polsce i strefie euro, *NBP, Materiay i Studia*, **227**, Warszawa, 2008.

[10] Skrzypczynski, P.: Metody spektralne w analizie cyklu koniunkturalnego gospodarki polskiej, *NBP, Materialy i Studia*, **252**, Warszawa, 2010.

[11] The World Bank, *http://data.worldbank.org/indicator*, 2012.

Agent based simulation of the selected energy commodity market

Marek Zachara[1], Dariusz Pałka[2], Ewa Majchrzyk-Zachara[3]

Abstract. A traditional macroeconomic approach to the analysis of commodity prices is usually based on the aggregate supply and demand as well as the abstract price elasticity of the demand. Unfortunately, this elasticity can be determined a posteriori (i.e. before a certain stimulus is experienced), which makes it difficult to predict even the most general market changes. The paper presents a simulation model developed by authors that incorporates major elements of the market, including customers, suppliers, storage, etc. The proposed simulation has an important advantage over a traditional approach, as it allows to predict the market behaviour on the basis of the model parameters which are identifiable and measurable a priori, i.e. infrastructure parameters and market participants behavioural parameters. that are acquirable with the use of existing sociological tools and models. The simulation model assumes the independence of action of every entity (agent) with certain regulatory and physical constrains, such as the supply routes capacity, environmental conditions, enforced minimal reserves, the individual customer priority over commercial entities, etc. The primary benefit of the model and the simulation environment is the possibility of observing the market under various hypothetical scenarios.

Keywords: agent simulation, energy market, consumer behaviour.

JEL classification: C63
AMS classification: 91B26, 68U20

1 Introduction

Forecasting and estimation of market behaviour under certain economic conditions has been a vital desire of many market analysts. For obvious reasons, however, it is not possible to conclude a scientific experiment in a controlled environment that would either prove or falsify certain economic theories. Therefore, the predictions and market analysis is often performed on the basis of similarities to the conditions observed earlier or the 'intuition' of the analyst. Now, however, there exists another option for predicting the behaviour of complex systems, namely, a simulation. For human behaviour modelling, agent-based [7] scenarios are often successfully employed [11], [1], [9].

In this paper a similar, agent-based approach inspired by [1] is utilised to simulate a national coal market. Specifically, the JADE [4] platform is used as a basis for the simulation. The agents implement basic economic behaviour, i.e. maximizing individual profits or minimizing the costs [3], [5]. Consumer models also include coal consumption rate which is calculated based on current external temperature and consumer's individual parameters (i.e. house heating requirements, preferred temperature).

The initial sections of the paper describe sellers and buyers models used for the simulation. The models were initially developed by the authors for the purpose of price dispersion analysis [12], and has been adjusted for the presented task. This paper also includes a new proposed model of a trader (middle man). Later on, selected simulation results obtained with the models described are presented. As can be seen, the collective intelligence represented by a set of virtual agents yields the results which are consistent with expectations, and can also be used to assess how the market could behave under certain conditions.

[1]AGH University of Science and Technology, Department of Applied Computer Science, Poland, mzachara@agh.edu.pl
[2]Pedagogical University of Cracow, Department of Computer Science, Poland, dpalka@ap.krakow.pl
[3]Not associated with a research institution, ewa@zachara.name

2 The models of market participants

2.1 Consumers

For the purpose of the simulation, a single type of customer is considered. It is assumed that the customer utilises the coal for heating. Therefore, their daily usage will mainly depend on the environmental factors (the temperature outside, the size of the house). The model of the consumer (C) is a simplified version of the model proposed in [12] for the analysis of the price dispersion [2] and consists of the following elements:

$$C\left\{L, S, K, E(ps, hs, tc, ex)\right\} \tag{1}$$

Where:

- L is the location of the customer
- S is their current stock of coal
- K is the knowledge possessed by the customer (about traders' offers)
- E is their personality.

E_{ps} denotes their sensitivity to the coal price, E_{hs} is the unified equivalent of their house heating requirements, while E_{tc} is the preferred temperature inside. E_{ex} denotes certain customers' willingness to switch to other sources of heating. Although energy prices are somehow correlated [8], temporary differences can often trigger customers' decisions to switch to a different source. The usage of the coal is therefore calculated as follows:

$$\Delta S = \begin{cases} -e_f \cdot E_{hs} \cdot (E_{tc} - t_{curr}) & if \quad E_{tc} > t_{curr} \\ 0 & if \quad E_{tc} \leq t_{curr} \end{cases} \tag{2}$$

Where:

- ΔS is the change in consumer's coal stock
- e_f is the heating efficiency of the coal
- t_{curr} is the current temperature outside

Each customer evaluates their stock of coal (S) daily and with a probability equal to the stock 'emptiness' (S_a) investigates the market and gathers knowledge (K) by polling the available traders for price offers. Upon receiving the offers, the customer then decides to replenish their stock of coal with the following probability:

$$\begin{aligned} H &= E_{ps} \cdot \left(a - b \cdot \tfrac{P_{tp}}{P_{atp}} - c \cdot S_a\right) \\ R_{cp} &= (1 - d \cdot E_{ps}) + sign(H) \cdot H^2 \end{aligned} \tag{3}$$

Where:

- R_{cp} is the probability of purchasing coal from a trader (the value is limited to (0-1) range)
- S_a is the percentage of available (empty) stock space the customer currently has
- P_{tp} is the current best price of coal available from the traders
- P_{atp} is the averaged (over time) price of the coal purchased by the consumer
- E_{ps} is the customer's personality (sensitivity to the price of coal)

a, b, c and d are the parameters which are adjusted to provide a reasonable environment: i.e. a customer with an average E_{ps} will have R_{cp} close to 100% when his stock (S_{ta}) reaches 50%, while an extreme saver will only have 75% chance of deciding to buy coal even if his stock is empty if prices are 20% above the average.

2.2 Traders

Traders are the entities that purchase coal in bulk from producers and sell it to consumers. They tend to have a relatively large stock of coal and adjust rather quickly to the market conditions. Traders have to

manage two sides of the transaction, i.e. purchasing the commodity and deciding on the price at which it will be offered to customers. Traders are modelled as follows:

$$T\{L, S, P, E(ta, tb))\} \tag{4}$$

Where:

- L is the location of the trader
- S is their current stock of coal
- P is the price offered
- E is their personality, denoting 'greediness' in terms of sensitivity to prices offered by producers

The purchase is determined by the current price of coal (compared to the long time average), and the current state of the trader's stock (he will be more pressed to buy if his current stock level is low). The probability and the amount of the purchase is defined as follows:

$$\begin{aligned} R_{tp} &= S_{ta} + t_a \cdot \left(100\% - \tfrac{P_{vp}}{P_{avp}}\right) \\ Q_{tp} &= S_{tt} \cdot \left(0.5 \cdot S_{ta} + t_b \cdot \left(100\% - \tfrac{P_{vp}}{P_{avp}}\right)\right) \end{aligned} \tag{5}$$

Where:

- R_{tp} is the probability of purchasing a delivery of coal by the trader (limited to (0-1) range)
- Q_{tp} is the amount (quantity) of coal requested
- S_{ta} is the percentage of the available (empty) stock space the trader currently has
- S_{tt} is the total size of the traders' stock space
- P_{vp} is the current best price of coal available from the producers
- P_{avp} is the averaged (over time) price of coal purchased by the trader
- t_a and t_b are parameters denoting particular trader's sensitivity to the purchase price

The traders conduct all transactions (provided they do have the stock to trade) if a consumer requests a purchase at the price offered by the trader. Each trader, however, can individually set their offer price. It is assumed that each trader's single and primary goal is to maximize the profits. Since the trader does not know how the market will respond to the offer in price change, during each evaluation cycle they take a risk and change their offered price taking into account the history. If the last change (e.g. decrease in the offered price) resulted in an increased overall profit for the trader, the trader will continue to change the price in the same direction (e.g. decrease the price further). If the last change (e.g. a decrease in the offered price) resulted in a decreased overall profit for the trader, the trader will change the price in the other direction (e.g. this time increase the price). In this case, the change is smaller in value (i.e. half of the previous change) to prevent artificial oscillations.

$$P(t_0) = \begin{cases} P(t_{-1}) + 1.5 * (P(t_{-1}) - P(t_{-2})) & \textit{if profit increased} \\ P(t_{-1}) - 0.5 * (P(t_{-1}) - P(t_{-2})) & \textit{if profit decreased} \end{cases} \tag{6}$$

2.3 Producers

Producers are the entities delivering the commodity (i.e. coal) into the market. For the purpose of the simulation, only one producer was used, since with such simple constrains (a single type of commodity, price based purchase decisions only), the producers end up following each other's pattern. Also, the external sources (importers) are not distinguished. In both cases, under a short-term scenario, the cost of producing (i.e. making available) a single unit of coal can be deemed constant. For internal producers this depends mostly on the cost of work. For importers it depends on the international prices (mostly governed as futures contracts) [6], [10]. It is also assumed that the production price is constant over the course of simulation. The producer is represented by the following model:

$$V\{S, P, G\} \tag{7}$$

Where:

S is the producer's current stock of coal

P is the price offered

G is their capacity to produce (units/time)

As far as determining the offered price is concerned, the producers follow the same pattern as traders, i.e. they adjust the offered price on the basis of the results from a previous change, as described in (6). The ultimate goal of the producer is to maximize the profit.

3 Initial model validation

In order to assess the model, it would be prudent to see how it behaves under the so called 'normal' conditions. For that purpose a simulating environment has been set up as follows:

the number of customers:	1000
the number of traders:	4
the time of the simulation:	5 years (1825 days)

Customers' and traders' personalities ($C.E$ and $T.E$) were randomized (evenly distributed within 0-100% range for each parameter). Some of the results of the simulation can be seen in Fig. 1 and Fig. 2. The first one (Fig. 1) illustrates the market from the producer's perspective. As can be seen, after the initial period of stabilizing the market, the prices offered and the producer's stock of coal change periodically with the demand driven by the change of the temperature. (The temperature is an actual historical average temperature in Poland retrieved from a weather web site).

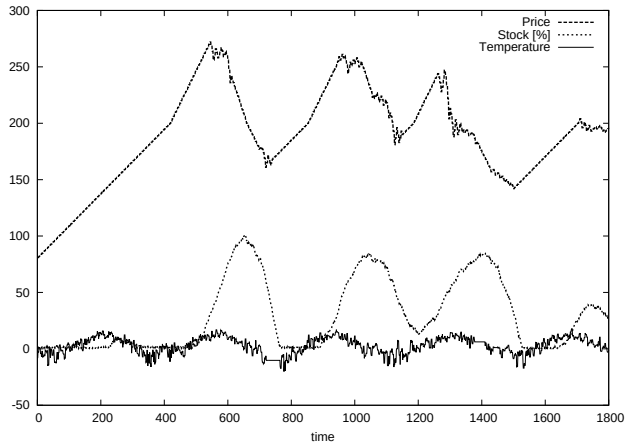


Figure 1 The price and stock of the producer compared to current outside temperature levels

Also, the behaviour of some consumers was illustrated in Fig. 2. Two specific customers were selected from the group of 1000: one with average parameters, and one with extremely 'greedy' personality (which means he would rather froze than pay a high price for coal).

These results indicate that the model, regardless of its simplicity, seems to work well and allows for reproducing the environment close enough to the real world. The results achieved are in line with what would be expected in real, unregulated market.

4 Hypothetical scenarios

In this section the results of some hypothetical scenarios will be given. These have a two-fold purpose. First, they allow for further verification of the applicability of the proposed model. Second, assuming the model has proven to be adequately close to the expectations, it gives us an insight into scenarios that can happen, but have not been experienced in real world yet.

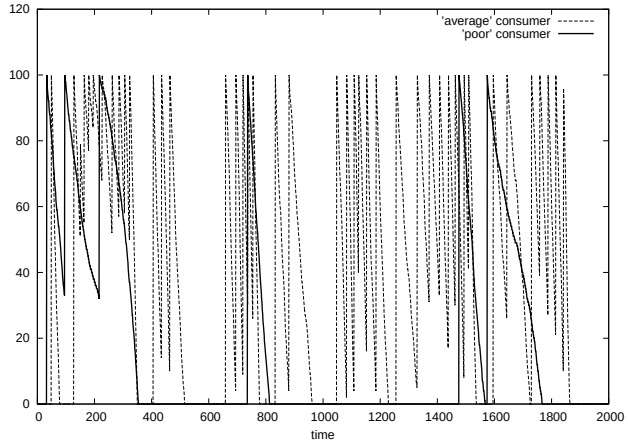


Figure 2 Coal stock (in %) of different types of consumers over time

4.1 No alternative, no limits

Under this scenario, there were no restrictions on the amount of money customers have to spend on purchases, nor do they have an alternative heating to switch to. They still, however, exhibit greedy behaviour - the probability of a purchase depends on the perceived attractiveness of the offer as in equation (3). The observed outcome is a steady rise in prices - even though customers try to purchase coal as cheaply as possible (and the price varies over time), the underlying condition of scarce commodity drives the prices steadily higher.

On the other hand, there is a single condition that may turn the scenario upside down. If the supply of coal is higher than the average demand, it drives the prices steadily down, down to the level of production costs. With no possibility to reduce the output, the prices are driven even further down, almost to zero. This is somehow a confirmation of the principle of the economy, i.e. that the rules of economic behaviour apply only to scarce resources.

4.2 Mass consumer flight

Another scenario that has been tested is a significant number of customers leaving the market (e.g. because of switching to another energy source). The results of this scenario are illustrated in Fig. 3. In this scenario, around half of the customers gradually leave between the t+1000 and t+1500 of the simulation. As can be seen, this results in a massive drop in prices and surge in producer's stock that is not being picked up by the traders. In Fig. 3, the producer's price *Price2* and their stock of coal *Stock2* is presented against a baseline scenario (*Price1* and *Stock1* respectively).

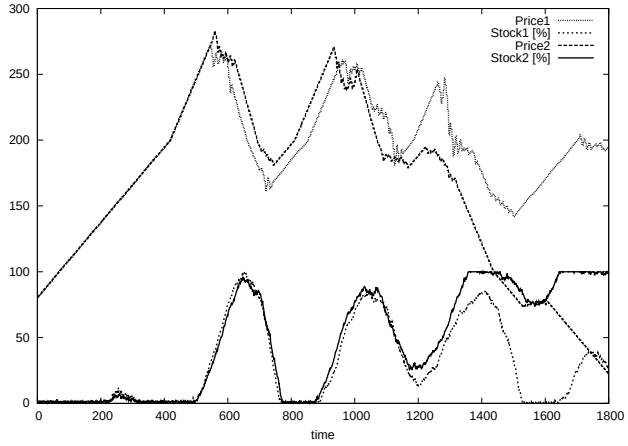


Figure 3 The results of a substantial number of customers leaving the market

4.3 Change in available stock sizes

Another scenario that has been tested is a change in the size of stock available to traders and producers. Surprisingly, a significant increase in the stock size (beyond a yearly turnover) resulted in the increased price volatility. This was likely due to a much lower number of individual transactions that do not let the traders 'keep in touch' with the market changes.

5 Conclusions

In this paper the use of agent-based models for simulating the behaviour of real markets is proposed. As has been proven, such environments behave in a way similar to the real world, provided that the models of the participants include key aspects of their behaviour. It has been demonstrated that relatively few constrains and behavioural scenarios are needed to implement a functional market.

Certainly, the primary requirement (and a basic validation of the models) is the similarity in behaviour of the artificial environment to the observations from the real world under typical circumstances. Yet, the real strength of the models is the possibility to observe the environment under certain set of unusual conditions.

Due to limited space, these hypothetical, unusual conditions have been discussed very briefly, yet they give an interesting insight into possible outcomes and prompt further research into the subject. Fortunately, the implemented simulation environment allows for relatively easy enhancement of the agents, which will allow for testing of more complex and sophisticated scenarios.

References

[1] Bonabeau E.: Agent-based modeling: Methods and techniques for simulating human systems, In: *Proceedings of the National Academy of Sciences*, **3(99)**, 2002, 7280–7287.

[2] Baye, M.R., Morgan, J., Scholten, P.: Information, Search, and Price Dispersion. In: *T. Hendershott, Editor, Handbook on Economics and Information Systems*, Elsevier, New York, 2006, 323–375.

[3] Begg, D., Begg, D.K.H., Fischer, S., Dornbusch, R.: *Economics*. McGraw-Hill, 2005.

[4] Bellifemine, F., Caire, G., Pogg, A., Rimassa, G.: JADE - A White Paper. In: *Telecom Italia Journal: EXP - in search of innovation*, **08**, Rome, 2003.

[5] Fishbein M.: Attitudes and Prediction of Behaviour. In: *Readings in Attitude Theory and Measurement*, John Wiley, New York, 1967, 477-492.

[6] Li, R., Joyeux, R., Ripple, R.D.: International steam coal market integration. *The Energy Journal*, **31**, 2010, 181–202.

[7] Lin, H.: *Architectural Design of Multi-Agents System*, Information Science Reference, 2007.

[8] Papiez, M., Smiech, S.: The analysis of relations between primary fuel prices on the European market in the period 2001-2011. In: *Rynek Energii*, **5(96)**, 2011, 139–144.

[9] Raberto M., Cincotti S., Focardi S.M., Marchesi M.: Agent-based simulation of a financial market. In: *Physica A: Statistical Mechanics and its Applications*, Elsevier, New York, 2001, 319–327.

[10] Smiech, S., Papiez, M.: Short-term predictions of import prices of steam coal in Poland. In: *Rynek Energii*, **2(99)**, 2012, 148–153.

[11] Was J., Kulakowski K.: Agent-based approach in evacuation modeling. In: *Agent and Multi-agent Systems: Technologies and Applications - Lecture Notes in Computer Science*, Springer-Verlag, 2010, 325–330.

[12] Zachara M., Piskor-Ignatowicz C.: Agent Based Simulation of Customers Behavior for the Purpose of Price Distribution Estimation. In: *Computational Collective Intelligence. Technologies and Applications - LNCS*, Springer, 2011, 211–220.

The usage of linear programming to constructing the ecologic-economical model for the industrial company profit optimization

František Zapletal[1], Radek Němec[2]

Abstract. The aim of this article is the application of the industrial company profit optimization. There is used the linear programming model to determine the optimal production structure. The built model is ecologic-economical – it involves indicators which have the impact on the environment, precisely the gas wastes. This is the reason why the model is constructed for the companies which have the business based on the heavy industry. Only such firms have to involve the ecological costs in their calculations.
The first part optimizes only the company profit from the sales which are put into the objective function. The ecological aspect of this part is given only by model constraints and the only source of the profit is the sold company products.
The second section of the article extends the model by the second objective function. It supposes that the industrial company profit does not consist only of the product sales, but it involves the profit generated by the sales of the emission permits too.
The model is tested on the historical but real data of one industrial company. Conclusions of the paper demonstrate the applicability of the model and possibilities of its extensions.

Keywords: optimization, linear programming, ecologic-economical model, emission permits.

JEL Classification: C61, D22

AMS Classification: 90C05

1 Introduction and objectives

Nowadays in crisis of the economy it is extremely important that the companies have to behave utmost rationally and they must fully exploit their potential. On the other hand all economic subjects are pushed to the environmental responsibility, which generally goes against their economic interests.

This paper deals with the design of the rationalization of the heavy-industry company production structure, which is limited not only by specific technological constraints, but also by the emission limits imposed by the state. The aim of this paper is to set up two optimizing models. The first model will assume that the company does not have an access to the marketplace with the emission allowances and the second one, conversely, will take into account also the profits from the trading with the emission permits. Impacts and efficiency of trading with emission permits are very often analyzed, but mostly from the macroeconomic point of view ([1], [3]). This paper deals with this problem at the firm level in the second constructed model.

For the large industrial companies in particular, the ecological costs are very important component of the total costs, however, not enough attention is paid to them.

The constructed models should serve as the decision support base for authorized subjects and they should enable:

- to analyze and assess the ecologic and economic requirements of each product;
- to optimize the amount of production with regard to the ecologic demands;
- to improve the allocation of investments into the technologies that would provide lower ecologic costs;

[1]VŠB - Technical University of Ostrava, Department of System Engineering and informatics, Sokolská 33 Ostrava, Frantisek.zapletal.st@vsb.cz.
[2]VŠB - Technical University of Ostrava, Department of System Engineering and informatics, Sokolská 33 Ostrava, radek.nemec@vsb.cz.

- to find the trade-off between the ecologic costs (emission limits) and the profits from selling the products in the first model and to extend this solution by the possibility of trading with the emission permits in the second model.

2 Data and methodology

The model will be tested on the real data of certain heavy-industrial company that prefers to remain anonymous in this article. Provided data will be further modified at the request of this company. The data about emissions were obtained from the databases of the Czech Environmental Information Agency (CENIA) and the Ministry of the Environment of the Czech Republic.

The optimization method that is used in both models is the linear programming. This method of the operation research is a very effective instrument for the processes optimization and it is used in wide range of disciplines where it is need to optimize – from generally used problems like project scheduling [7] to many specialized issues like planning the e-learning courses [2]. Linear programming was chosen, because all the constraints and even the objective function are described by the linear equations and inequations. A lot of algorithms can be used to find the optimal solution of the linear programming problem; the simplex method [4] is used in this paper. Optimization was implemented in Microsoft Excel 2007 which contains a built-in solver tool that uses a simplex method [8].

3 The profit optimization of industrial company with no emission permits trading

For the purposes of this model we assume, that the company makes a profit only by selling its products. Thus, the company solves a simple profit maximization problem (or the losses minimization). The ecologic aspect cannot be the source of profits, but it can induce additional costs. These costs would arise if the company exceeds the emission limits. In such a case, the company would have to pay the fines. Therefore, the emission limits have to be taken into account together with the remaining constraints of the production.

3.1 The structure of the model

The objective function consists only of the profits from selling the production. The model condition is that whole production is sold and thus brings the expected profit to the company [9]. The objective function is described by expression (1).

$$\sum_{i=1}^{s} x_i \cdot \pi_i \rightarrow max \qquad (1)$$

where x_i is the amount of the i-th product and π_i is the profit generated by one unit of the i-th product. The company makes s products in total.

In every production planning problem of the linear programming, the non-negativity constraints must be fulfilled (the produced amounts cannot be negative); this is given by inequation (2). Then, the available amount of limited resources needed for the production has to be taken into account (e.g. the capacity of the machines, stocks, labour etc.). These constraints are given by inequation (3). Other constraints are given by actual demand for particular products of the company inequation (4). In the short range, the contracts are already made with the customers, so it is not possible to decrease the produced amounts under these values even in the case when the would allow the company to reach higher profits. The last constraint given by inequation (5) represents the emission limits; p_{ji} = amount of the j-th pollutant emitted in making one unit of the i-th product, l_j = the emission limit for the j-th pollutant)

$$x_i \geq 0 \qquad (2)$$

$$x_i \leq capacity \qquad (3)$$

$$x_i \geq demand \qquad (4)$$

$$\sum_{i=1}^{s} p_{ji} \cdot x_i \leq l_j \qquad (5)$$

3.2 The application of the model on the real data

The modeled company produces 8 products and each of them is sold for the different price (Table 1). They are produced only 4 emissions just for the model simplification (Table 2). The production is limited by the capaci-

ties, which are described in table 4, where are actual demanded amounts for each product. This demand must be covered by production within a set time, because of the made contracts with the customers of the company.

Product (i)		Profit
1	Raw iron	2.2028
2	Brams	1.4160
3	Ingots	3.8846
4	Blocks	0.6200
5	Iron plates 4.5	0.1145
6	Iron plates 3.5	0.8054
7	Profiles	2.3938
8	Cut shapes	0.1628

Table 2 Profit from unit of i-th product [in thousands of CZK]

j	Units	l_j
SO2	kg	2000000
NOX	kg	2500000
CO	kg	70000000
Ash	kg	900000

Table 1 List of emissions and their limits

		Product (i)							
		1	2	3	4	5	6	7	8
SO2	kg	0.5787	0.1853	0.1329	0.1498	1.273	1.4193	1.1093	1.1117
NOX	kg	0.1983	0.7231	0.5482	0.5829	1.4728	1.6574	1.3968	1.5003
CO2	kg	10.7263	58.9362	45.7492	50.0721	0.5938	0.7334	0.6932	0.6777
Ash	kg	0.5929	0.4302	0.4284	0.4401	0.0529	0.0601	0.0573	0.0589

Table 3 Amount of j-th emission produced by 1 ton of i-th product

	Product (i)							
	1	2	3	4	5	6	7	8
Capacity	1000000	900000	40000	100000	100000	700000	90000	12000
Demand	651802	831602	38400	95382	50000	608600	90000	11800

Table 4 Capacity and actual demand of each product [tones]

$$2.2028x_1 + 1.4160x_2 + 3.8846x_3 + 0.6200x_4 + 0.1145x_5 + 0.8054x_6 + 2.3938x_7 + 0.1628x_8 \rightarrow max$$

Objective function 1 Objective function of the model involving allowance trading

$$x_1 \geq 651802;\ x_2 \geq 831602;\ x_3 \geq 38400;\ x_4 \geq 95382;\ x_5 \geq 50000;\ x_6 \geq 608600;\ x_7 \geq 90000;$$
$$x_8 \geq 11800.$$

Inequation array 1 Constraints given by demand

$$x_1 \leq 1000000;\ x_2 \leq 900000;\ x_3 \leq 40000;\ x_4 \leq 100000;\ x_5 \leq 100000;\ x_6 \leq 700000;\ x_7 \leq 90000;$$
$$x_8 \leq 12000.$$

Inequation array 2 Constraints given by capacity

$$0.579x_1 + 0.185x_2 + 0.133x_3 + 0.15x_4 + 1.273x_5 + 1.419x_6 + 1.109x_7 + 1.1117x_8 \leq 2000000$$

$$0.198x_1 + 0.723x_2 + 0.548x_3 + 0.583x_4 + 1.473x_5 + 1.657x_6 + 1.397x_7 + 1.5003x_8 \leq 2000000$$

$$10.73x_1 + 58{,}94x_2 + 45{,}75x_3 + 50.07x_4 + 0{,}594x_5 + 0.733x_6 + 0.693x_7 + 0.6777x_8 \leq 70000000$$

$$0.593x_1 + 0.43x_2 + 0.428x_3 + 0.44x_4 + 0.053x_5 + 0.06x_6 + 0.057x_7 + 0.0589x_8 \leq 900000$$

Inequation array 3 Constraints given by emissions limits

Results

The results of the optimization (Table 5) show, that the company can reach the higher profits by increasing in production. Concretely it should be increased the produced amount of the raw iron (by 11.93 %), ingots (by 4.17 %) and iron plates 3.5 (by 15 %). These changes would lead to growth of profits by 7.1 % (it means more than 251 million per year).

	x1	x2	x3	x4	x5	x6	x7	x8	Profit
Before opt.	651802	831602	38400	95382	50000	608600	90000	11800	3534897
After opt.	729573	831602	40000	95382	50000	700000	90000	11800	3786044
Difference [1000CZK]	77771	0	1600	0	0	91400	0	0	251147,34
Difference [%]	11.93 %	0 %	4.17%	0 %	0 %	15.02%	0 %	0 %	7.10%

Table 5 Results of optimization (in thousands of CZK and percents)

4 The profit optimization of industrial company with the emission permits trading

In this part of the article the model is extended by the possibility of firm to deal with the allowances in the emission permits market. The profits maximization still remains the only one goal of the company. But the profits consist of two components in this case – the profit from selling the production and the profit from selling emission allowances. A both these components are conflicting – when the production quantity increases, the profits from selling this production will increase too. On the other hand it would be produced the larger amount of the emissions whose allowances could be sold by the company and therefore the profits from the selling allowances decreases. The change of the production structure is effective when the growth of one component of the profit exceeds the decrease of another component.

The objective function (1) from the previous model remains and it is complemented by another one, which maximize the profits from selling the allowances (6). The supposition is that the company is able to sell all tradable emissions, which it did not exploit with regard to the given emission limits (3).

4.1 The structure of the model

$$\sum_{j=1}^{t}\left(l_j - \sum_{i=1}^{s} p_{ji} \cdot x_i\right) \cdot r_j \rightarrow max \qquad (6)$$

, where r_j is market price of the one permit (one permit = one ton of the emission) of the j-th pollutant (used units must be same like the units of profit from production selling).

Because of the same units and same variables which are used in both objective functions, they can be aggregated to the one objective function (7).

$$\sum_{i=1}^{s} x_i \cdot \pi_i + \sum_{j=1}^{t}\left(l_j - \sum_{i=1}^{s} p_{ji} \cdot x_i\right) \cdot r_j \rightarrow max \qquad (7)$$

Expression (7) can be further suitably arranged by expressing the variable x. The final form is in expression (8).

$$\sum_{i=1}^{s} x_i \left(\pi_i - \sum_{j=1}^{t} p_{ji} \cdot r_j\right) + \sum_{j=1}^{t} l_j \cdot r_j \rightarrow max \quad (8)$$

The constraints of the model remain the same like in the case when the company could not realize allowances selling (Inequations from (2) to (5)).

4.2 The application of the model on the real data

In the Czech Republic (or more precisely E.U.) there are currently only allowances for CO_2 traded [6]. Average market price of these allowances is CZK 201.79 per allowance that is per 1 ton of CO_2 [5]. Objective function of the model, which encompasses trading only one kind of emission, is shown by objective function 2. All other inputs are taken from the previous model (i.e. inequation arrays 1, 2 and 3)

$$2.20066x_1 + 1.40412x_2 + 3.87542x_3 + 0.60996x_4 + 0.11441x_5 + 0.80528x_6 + 2.39366x_7 + 0.16264x_8 + \\ +14700 \rightarrow max$$

Objective function 2 Objective function of the model involving allowance trading

	x1	x2	x3	x4	x5	x6	x7	x8	Profit
Before opt.	729573	831602	40000	95382	50000	700000	90000	11800	3786044
After opt.	729573	831602	40000	95382	50000	700000	90000	11800	3787239

Table 6 Results of optimization (in thousands of CZK and percents)

Profit from permits		Profit from products	
0	0.000%	3786044	100%
1227	0.032%	3786044	99.968%

Table 7 Structure of optimized profit

Tables 6 and 7 show optimization results and equate them with the initial state – i.e. the result of the first model (see table 5). They show the optimized production without the possibility of trading allowances. Results show that the possibility of trading emission allowances for CO_2 will lead to increase in total profit of the firm by nearly CZK 1.2 million. The structure of the production will not change. To change optimal amount of production of each product there should be an increase in the price of allowances or there need to be also other emissions than carbon dioxide started to be traded on the European market (in the USA there is also nitrogen oxide and sulfur dioxide traded).

Now there will also be made an optimization of the same model with the same data, only the constraints given by demand are removed. Results could be vital for production planning in the future, where the still running customer contracts will come to the end and there will be an opportunity to downsize production volumes. It is to be emphasized that the model counts on constant coefficients and does not capture their (real or potential) evolution in time.

	x1	x2	x3	x4	x5	x6	x7	x8	Profit
Before opt.	729573	831602	40000	95382	50000	700000	90000	11800	3787239
After opt.	1000000	564242	40000	0	0	700000	90000	0	3941344
Difference [1000CZK]	270427	-267360	0	-95382	-50000	0	0	-11800	154105
Difference [%]	37.07%	-32.15%	0%	-100%	-100%	0%	0%	-100%	4.07%

Table 8 Results of optimization (in thousands of CZK and percents)

Profit from permits		Profit from products	
1227	0.032%	3786044	99.968%
4959	0.126%	3936385	99.874%

Table 9 Structure of optimized profit

From results in tables 8 and 9 stems the fact that if the firm is not bottom limited by contracted demand, the production structure would change. Raw steel production would rise by more than 37%, amount of produced brams would vice-versa fall by more than 32% and the production of 3 products – blocks, plates 3.5 and cut shapes even stops. The new profit would rise by approximately 4% compared to the original. Here is to emphasize that although the model can determine optimal amount of production of each product in a way that the firm would generate maximal profit but it does not handle output for these products. So the computed profit is achievable only if the firm can manage to sell all of its production.

5 Conclusion

The aim of the paper was to create proposal of an economic-ecological model, which would optimize production of an industrial company with respect to a maximization of the profit from selling its production and profit from allowances selling. Through the application of the model on the real data was discovered the fact, that if the company enters the emission allowances market the total profit could be raised, nevertheless this appraisal would be compared to the profit from selling the production very low and will not have the impact on the production volumes. The reason is relatively low actual price of traded allowances and also a fact that in the European Union there is only one kind of emission traded, which is carbon dioxide. If there would be a change in conditions for trading emission allowances (in a way of extending the market for other emissions or appraisal of prices) the model could be able to flexibly react and the share of profit from allowances trade on total profit would rise.

The model could be further extended. It is possible to alter it for conditions of specific company or extend the model structure, so the scope would be wider. One of proposal is that the model could be enriched by analysis of sensitivity on the changes in price coefficients, especially prices of emission allowances and unitary profit for products.

Acknowledgements

This paper was made under financial support of Student Grant Competition, research project SP2012/184 "Analýza charakteristik modelování databázových schémat datového skladu s důrazem na agilní přístup k vývoji Business Intelligence systému."

References

[1] Ben-David, S.; Brookshire D. S.; Stuart B.; McKee M.; Schmidt Ch.: Heterogeneity, Irreversible Production Choices, and Efficiency in Emission Permit Markets. *Journal of environmental economics and management* **38 (**1999), 176-194.

[2] Fernandez, S.; Borrajo, D.: Using linear programming to solve clustered oversubscription planning problems for designing e-courses. *Expert systems with applications* **5** (2012), 5178-5188.

[3] Gersbach, H.; Winkler, R.: International emission permit markets with refunding. *European Economic Re view* **55 (**2011), 759-773.

[4] Hillier, F. and Lieberman G. J.: *Introduction to operation research*. 9th ed. 1047 p. McGraw-Hill Higher Education, New York, 2010.

[5] Ministry of the Environment of the Czech Republic. *Průměrná cena emisní povolenky k 28. únoru 2012* [on line], [cit. 2012-05-02]. Available from WWW:<http://www.mzp.cz/C1257458002F0DC7/cz/cena_emisni povolenky/$FILE/OEOK-Prumerna_cena_emisni_povolenky_20120228.pdf>.

[6] Ministry of the Environment of the Czech Republic. *Národní alokační plan České republiky na roky 2008-2012* [online], [cit. 2012-05-02]. Available from WWW: <http://www.mzp.cz/C1257458002F0DC7/cz/narodni_alokacni_plan/$FILE/OZK-NAP_2-20081008.pdf>.

[7] Szmerekovsky, J. G.; Venkateshan, P. : An integer programming formulation for the project scheduling problem with irregular time-cost tradeoffs. *Computers & Operations research* **7** (2012), 1402-1410.

[8] Taha, A.: *Operation Research: An introduction.* 9th ed. 832 p. Pearson, New Jersey, 2011.

[9] Zapletal, F.: *Creation Of the Ecologic-Economical Model for an Industrial Company*: diploma thesis. Ostrava: VŠB – Technical University of Ostrava, Faculty of Economics, 2011. 62 p. Thesis supervisor: Jiří Chuchro.

Constructing business cycle regime switching model for Czech economy

Jana Závacká[1]

Abstract. This paper is oriented at modeling cyclical behavior of macroeconomic activity in the Czech Republic. It reacts on the temporal critique of assumptions that there are linear relationships between main macroeconomic variables while using vector autoregressive (VAR) or structural vector autoregressive (SVAR) approach. The model constructed here is based on vector autoregressive approach but it allows the changes in relationships by regime switching. The switch is made according to the business environment, more accurately according to phases of expansion or contraction in the economy. Considering that turning points in the economic activity could not be necessary determined only by exceeding fixed values of observed variables, regime switching model is made by using dummy variables in the model. The turning points' dating is adopted from OECD turning points dating for Czech economy. By distinguishing of the environment conditions in the model the different relationships between variables are investigated. As an endogenous variable in the model an indicator of confidence of firms is added, as a possible source of information about future development of economy.

Keywords: regime switching, indicator of confidence, business cycle, nonlinear model.

JEL Classification: C51, E32
AMS Classification: 91B64

1 Introduction

The changes in economic activity in the past years and especially long-lasting unexpected slowdown confirmed the need of improvement in business cycle modeling. Vector autoregressive (VAR) models, still used for modeling economic activity, were already criticized by Hamilton [4] because of its linearity assumption. Hamilton [5] mainly points out that traditional VAR models are weak around turning points in economic activity. To overcome this problem a various nonlinear autoregressive models were proposed. Tong and Lim [6] demonstrated that partially linear functions in Threshold autoregressive (TAR) models could well capture the cyclical movement in economic activity. Hamilton [4] proposed a solution through Markow switching models. Dueker and Assenmacher-Wesche [3] constructed Qual VAR model with information about business cycle movement as endogenous variable in VAR. Unfortunately the suitable way how to predict behavior of economic activity and especially its turning points was not very well described yet.

With the effort of investigation of modeling economic activity the aim of this paper is to propose and construct autoregressive model for the Czech economy, where the information about business cycle movement will be considered. Firstly, we want to find out, if the information about the business cycle could improve modeling economic activity around turning points. Secondly, by imposing the information about the economic environment, we want to find out if the modeling of behavior of main macroeconomic indicators could vary with the phase of business cycle. If the modeling of behavior of economic activity would be phase-sensitive, it could mean that not only the assumption of the linearity of functions in VAR models is weak but more the assumption that all the regressors in VAR models have constant and continuous effect in the time on the economic activity is weak. For imposing the business cycle information into the model the threshold autoregressive model is used. There is a regime switching between expansion and contraction throughout the business cycle. The thresholds are set to fixed turning points dated for Czech economy by OECD. The switch between two phases of business cycle is made by partially linear functions. For expansion as well as for contraction the vector autoregressive model with restrictions are estimated. The restrictions on parameters in both models are set to follow the main macroeconomic principles.

The paper is structured as follows. In section 2 the model is introduced. The estimation results of the model are to be found in section 3. Section 4 concludes.

[1]Department of Mathematical Methods in Economics, Faculty of Economics, VŠB - Technical University of Ostrava, Sokolská třída 33, 701 21 Ostrava, Czech Republic
jana.zavacka1@vsb.cz.

2 The model

The model is based on quarterly data from the Czech economy from 1996Q1-2011Q4. As endogenous variables of the model were chosen GDP at market prices (chain-linked volumes with reference year 2005) (HDP_t^{real}), the rate of unemployment (mi_t), harmonized index of consumer prices (with reference year 2005) ($HICP_t$), interest rate PRIBOR3m (ir_t) and indicator of confidence of firms (with reference year 2005) (ID_t^{firm}). Because of the small and open character of Czech economy the exogenous variable gross domestic product in Eurozone (chain-linked volumes with reference year 2005) (HDP_t^{EA}) is added to the model, too. We assume that for all these variables there exists their constant steady states values (or steady states constant rates of growth) and transform the data into the deviations from the steady states. In case of GDP, HICP, indicator of confidence of firms and GDP of EA we work with their year-on-year growth rates and in case of interest rate and rate of unemployment we work with their year's difference. After this transformation all the time series could be according to Dickey-Fuller test and Phillips-Perron test considered as stable. Additional information into the model is a chronology of turning points of economic activity in the Czech Republic dated by OECD: Trough 1994M1, Peak 1996M1, Trough 1996M11, Peak 1998M2, Trough 2000M1, Peak 2001M4, Trough 2003M5, Peak 2008M2, Trough 2009M4. We transform this chronology into the quarterly data by the rule that prevailing phase in the quarter defines the phase of the whole quarter. All the estimations are made in program Stata.

The decision of regime switching between expansion and contraction throughout the business cycle comes from the hypothesis that the relationships between variables as well as intensity how they can influence each other could vary through the cycle. For that reason every estimated parameter in the model is always related to variable and to the phase in which business cycle already is. The model is proposed in accordance with the logic of threshold autoregressive model presented in Arlt and Arltová [1]. However the thresholds are not set on fixed constants but are fixed on dated turning points. To incorporate this setting into the model the dummy variable θ_t is defined. $\theta_t = 1$ in case that economy is in expansion in time t, $\theta_t = 0$ in case of contraction. The model is then defined

$$\begin{aligned} y_t = \theta_{t-1} \cdot B_{t-1}^E \cdot y_{t-1} + \theta_{t-2} \cdot B_{t-2}^E \cdot y_{t-2} + (1-\theta_{t-1}) \cdot B_{t-1}^K \cdot y_{t-1} + (1-\theta_{t-2}) \cdot B_{t-2}^K \cdot y_{t-2} + \theta_{t-1} \cdot c_{t-1}^E \cdot \\ \cdot x_{t-1} + \theta_{t-2} \cdot c_{t-2}^E \cdot x_{t-2} + (1-\theta_{t-1}) \cdot c_{t-1}^K \cdot x_{t-1} + (1-\theta_{t-2}) \cdot c_{t-2}^K \cdot x_{t-2} + \varepsilon_t, \end{aligned} \tag{1}$$

where $y_t = \left(HDP_t^{real}, mi_t, HICP_t, ir_t, ID_t^{firm}\right)'$ denotes vector of endogenous variables at time *t*, $x_t = HDP_t^{EA}$ exogenous variable at time t, B_{t-I}^E, B_{t-I}^K matrices of parameters at time *t-i* in expansion, in contraction, respectively, c_{t-I}^E c_{t-I}^K vectors of parameters at time *t-i* in expansion, in contraction, respectively, and ε_t multivariate white noise. The order of the model was set to 2 to achieve stability in the model. The model could be understood as a threshold autoregressive model with many thresholds around turning points. Unfortunately we do not have enough data to estimate every partial linear function around turning points. For technical purposes we firstly estimate parameters in matrices B_{t-I}^E,C_{t-I}^E by OLS method from the sample of the data when all variables in the model are in expansion, e.g. the model

$$y_t = B_{t-1}^E \cdot y_{t-1} + B_{t-2}^E \cdot y_{t-2} + c_{t-1}^E \cdot x_{t-1} + c_{t-2}^E \cdot x_{t-2} + \varepsilon_t. \tag{2}$$

For the lack of the data we could only estimate B_{t-I}^K C_{t-I}^K from the sample, where all variables in time *t, t-1* are in contraction, e.g. from a part of the model

$$\begin{aligned} y_t = B_{t-1}^K \cdot y_{t-1} + \theta_{t-2} \cdot B_{t-2}^E \cdot y_{t-2} + (1-\theta_{t-2}) \cdot B_{t-2}^K \cdot y_{t-2} + c_{t-1}^K \cdot x_{t-1} + \theta_{t-2} \cdot c_{t-2}^E \cdot x_{t-2} + \\ + c_{t-1}^K \cdot x_{t-1} + (1-\theta_{t-2}) \cdot c_{t-2}^K \cdot x_{t-2} + \varepsilon_t. \end{aligned} \tag{3}$$

We use OLS estimate to and approximate this part of the model by the equation

$$y_t = B_{t-1}^K \cdot y_{t-1} + B_{t-2}^K \cdot y_{t-2} + c_{t-1}^K \cdot x_{t-1} + c_{t-2}^K \cdot x_{t-2} + \varepsilon_t. \tag{4}$$

If we consider, that the relationships between variables are developing, so in turning points could slowly transform, we assume that our approximation could not be much biased.

3 Estimation results

To achieve stability of the estimated model and omit the collinearity in the model we restricted the model into the order 2 in variables. Because the parameters of the estimated matrices were in many cases found after tests as

insignificant, we put into the matrices restrictions following macroeconomic principles. For all variables we allowed smoothing (in every equation every variable is estimated by its own lagged value). In equation for the rate of unemployment we according to the Okun's law dropped interest rate and HICP. Considering Phillips curve we concerned on the relationship between inflation and the rate of unemployment in the short-run. The interest rate is modeled with the respect to Taylor rule. Indicator of confidence of firms is allowed as a regressor in all equations. According to Barsky and Sims [2] the information about consumer confidence could be understood in two ways. Firstly, it is considered as an animal spirit, which could be the source of cyclic behavior. According to Barsky and Sims [2] it was not approved. Secondly, the confidence could contain some additional information about future development in economy, which is not a part of our data. This was confirmed (Barsky and Sims [2]). However we are working with the confidence of firms, not consumers (for indicator of confidence of firms longer time series available), there is a high correlation between both indicators and therefore we expect the indicator of confidence of firms could also include additional information about future development of economic activity. After setting these restrictions and eliminating insignificant coefficients we obtained following estimated equations of the model (p-value of significance of each regressor could be found under the estimated coefficient in the brackets):

$$HDP_t^{real} = \theta_{t-1} \cdot \left(\underset{(0.000)}{1.015} \cdot HDP_{t-1}^{real} + \underset{(0.000)}{0.192} \cdot ID_{t-1}^{firm} \right) + \theta_{t-2} \cdot \left(\underset{(0.000)}{0.625} \cdot HICP_{t-2} - \underset{(0.000)}{0.004} \cdot ir_{t-2} - \right.$$

$$\left. - \underset{(0.000)}{0.207} \cdot ID_{t-2}^{firm} \right) + (1 - \theta_{t-1}) \left(\underset{(0.000)}{0.631} \cdot HDP_{t-1}^{real} + \underset{(0.015)}{0.132} \cdot ID_{t-1}^{firm} \right) - (1 - \theta_{t-2}) \cdot \underset{(0.058)}{0.108} \cdot ID_{t-2}^{firm} -$$

$$- \theta_{t-1} \cdot \underset{(0.000)}{1.036} \cdot HDP_{t-1}^{EA} + \theta_{t-2} \cdot \underset{(0.005)}{0.554} \cdot HDP_{t-2}^{EA} + (1 - \theta_{t-1}) \cdot \underset{(0.018)}{0.298} \cdot HDP_{t-1}^{EA}$$

(5)

$$mi_t = \theta_{t-1} \cdot \left(\underset{(0.006)}{-3.42} \cdot HDP_{t-1}^{real} + \underset{(0.000)}{0.846} \cdot mi_{t-1} - \underset{(0.000)}{4.895} \cdot ID_{t-1}^{firm} \right) + \theta_{t-2} \cdot \underset{(0.001)}{2.815} \cdot ID_{t-2}^{firm} +$$

$$+ (1 - \theta_{t-1}) \cdot \left(\underset{(0.000)}{0.817} \cdot mi_{t-1} - \underset{(0.000)}{3.896} \cdot ID_{t-1}^{firm} \right)$$

(6)

$$HICP_t = \theta_{t-1} \cdot \left(\underset{(0.001)}{-0.192} \cdot HDP_{t-1}^{real} + \underset{(0.024)}{0.003} \cdot mi_{t-1} + \underset{(0.000)}{0.749} \cdot HICP_{t-1} - \underset{(0.001)}{0.057} ID_{t-1}^{firm} \right) +$$

$$+ (1 - \theta_{t-1}) \cdot \left(\underset{(0.000)}{0.752} \cdot HICP_{t-1} + \underset{(0.000)}{0.004} \cdot ir_{t-1} + \underset{(0.002)}{0.145} \cdot ID_{t-1}^{firm} \right) - (1 - \theta_{t-2}) \cdot \underset{(0.000)}{0.203} \cdot ID_{t-2}^{firm} +$$

$$+ \theta_{t-1} \cdot \underset{(0.000)}{0.887} \cdot HDP_{t-1}^{EA} + (1 - \theta_{t-1}) \cdot \underset{(0.000)}{0.587} \cdot HDP_{t-1}^{EA}$$

(7)

$$ir_t = \theta_{t-1} \cdot \left(\underset{(0.079)}{-6.614} \cdot HDP_{t-1}^{real} - \underset{(0.018)}{0.349} \cdot mi_{t-1} + \underset{(0.000)}{0.227} \cdot ir_{t-1} \right) + \theta_{t-2} \cdot \underset{(0.022)}{16.714} \cdot HICP_{t-2} +$$

$$+ (1 - \theta_{t-1}) \cdot \left(\underset{(0.152)}{-16.157} \cdot HICP_{t-1} + \underset{(0.018)}{0.369} \cdot ir_{t-1} + \underset{(0.032)}{19.25} \cdot ID_{t-1}^{firm} \right)$$

(8)

$$ID_t^{firm} = \theta_{t-1} \cdot \left(\underset{(0.000)}{-0.586} \cdot mi_{t-1} + \underset{(0.000)}{1.786} \cdot HICP_{t-1} - \underset{(0.000)}{0.479}\ ID_{t-1}^{firm} \right) + \theta_{t-2} \cdot \left(\underset{(0.000)}{-0.059} \cdot mi_{t-2} - \right.$$
$$\left. - \underset{(0.000)}{0.016} \cdot ir_{t-2} \right) + (1 - \theta_{t-1}) \cdot \underset{(0.000)}{0.964} \cdot ID_{t-1}^{firm} + (1 - \theta_{t-2}) \cdot$$
$$\left(\underset{(0.000)}{1.343} \cdot HICP_{t-2} - \underset{(0.002)}{0.606} \cdot ID_{t-2}^{firm} \right) - \theta_{t-2} \cdot \underset{(0.008)}{1.766} \cdot HDP_{t-2}^{EA} + (1 - \theta_{t-1}) \cdot \underset{(0.000)}{2.203} \cdot HDP_{t-1}^{EA} \tag{9}$$

Estimated model is stable and according to tests made in Stata (Lagrange multiplier, normality test) we could assume that ε_t is multivariate white noise. The majority of estimated parameters are significant on the 95% level of significance The first estimated part of the model, described by equation (2), has a value of log-likelihood funciton 248.389. All the estimations of intial equations were significant on the 95% level of significance, with R-square values 0.976 for real GDP, 0.942 for the rate of unemployment, 0.959 for HCIP, 0.47 for the interest rate and 0.804 for the confidence indicator of firms. The second estimated part of the model, estimated from the equation (4), has a value of log-likelihood function 118.854. All the estimated equations were also significant on the 95% level of significance, with R-square values 0.759 for the real GDP, 0.942 for the rate of unemployment, 0.964 for HICP, 0.588 for the interest rate, 0.86 for the indicator of confidence of firms.

It could be seen that in case of contraction the prediction is mainly based on its lagged values and indicator of confidence of firms. The indicator was founded as significant regressor (on the 95% level of significance) in nearly all equations. The estimated parameter 1,015 for lagged value of GDP in expansion and 0,631 in contraction in equation (5) is consistent with the idea that in expansion the growth of rate of GDP is growing and in contraction the speed is slowing down. The positive effect of the growth of HICP and the negative effect of the growth of interest rate on the growth of GDP in case of expansion is also favorable. On the Figure1 in Appendix the fit of the equation (5) could be seen. Unfortunately, it is observable that in turning points in the beginnings of contractions the proposed model does not fit well. Expected improvement in modeling economic activity in turning points was probably not well achieved. To state this more analysis should be made. For comparison VAR model without the business cycle information should be used, too. On the other side, the estimated equation confirmed that economic activity could be better modeled with respect to the business environment. The behavior of GDP is according to (5) well captured by the influence of the growth of the price level and the interest rate in the case of expansion, unlike in the case of contraction is better modeled according to the confidence indicator.

According to the estimated equation (6) the Phillips curve could be observed only in the phase of expansion. In equation (7) the price level is better modeled by the effect of the unemployment rate in the case of expansion and rather by the interest rate in the phase of contraction. The estimation of indicator of confidence (9) of firms is in the phase of expansion supported by all the main macroeconomics indicators but in the phase of contraction it is mainly dependent on its lagged values and GDP of Eurozone. In all these equations we can observe that the additional information about the economic environment changes structure of suitable model for tracing the behavior of chosen macroeconomic indicators.

The estimation of interest rate (8) could be definitely improved. Both R-squares of estimated parts of equations are low (0.47 for coefficients in expansion case, 0.588 in contraction). Some of the estimated coefficients are too high to be stable with high p-values. Unfortunately this equation after omitting insignificant variables does not fulfill according to tests the normality distribution of residuals. This could be much explained with the fact, that interest rate is under the control of the central bank. As a part of the proposed TAR model the fulfillment of the base assumptions was followed. This equation is definitely a weak part of the model.

4 Conclusion

In accordance with the goal of this paper the nonlinear autoregressive model for Czech economy was proposed. The main contribution of this model could be seen in discovering relationships between variables during different phases of the business cycle. It is already observed that the length of contraction is usually shorter than the length of expansion. Together with the phase-sensitive estimated parameters of the model we have to conclude that cyclical behavior is not a symmetric process. Not only the coefficients but also variables chosen for modeling economic activity differ with the business environment. From this perspective using traditional vector autoregressive models must not be the suitable way to follow the behavior of economic activity.

Unfortunately the partial linearization of the function probably did not improve the predictions around turning points. On the other side it could contribute to the predictions between them.

There are more conclusions which could be made according to the results of proposed model. The one is that with respect of the business environment we could investigate on which indicators we should be concentrated on if we want to predict or influence economy. From the results of estimated model it could be seen that in the phase of contraction the effort to fight against the economic slowdown should be mainly oriented on the increase of confidence of firms (perhaps even consumers, it was not studied). This should be also the main component in leading indicator for prediction of economic activity development. In case of expansion the leading indicators should be on the other hand more based on main macroeconomic indicators as HICP, unemployment rate, interest rate.

In addition, the indicator of confidence of firms was incorporated into the model, too. This is a new instrument, which could bear according to Barsky and Sims [2] additional information about future economic development. The significance of this indicator was approved by the estimations of the model, so the future development in this field could be beneficial, too.

The model is partially linear and regime switching is made through the dated chronology of turning points, posted by OECD. Because of this character any prediction or simulation, which should be run within this model, requires external simulation of this chronology. Nonetheless, the chronology posted by OECD was estimated from the basic macroeconomic indicators, probably some of them could be already in the model. The next development of this model could be therefore a construction of the mechanism of generating turning points in economy within the data from the model.

Acknowledgements

The paper was funded by Student Grant Competition of VŠB-TU Ostrava within the project SP2012/62.

References

[1] Arlt, J., and Arltová, M.: *Ekonomické časové řady*. Professional Publishing, Praha, 2009.
[2] Barsky, R. and Sims, E. R.: Information, Animal Spirits, and the Meaning of Innovations in Consumer Confidence. *NBER Working Papers* **15049** (2009).
[3] Dueker, M., and Assenmacher-Wesche, K.: Forecasting macro variables with a Qual VAR business cycle turning point index. *Applied Economics: Taylor and Francis Journals* **42** (2010), 2909–2920.
[4] Hamilton, J. D.: A new approach to the economic analysisof nonstationary time series and the business cycle. *Econometrica* **57** (1989), 357–384.
[5] Hamilton, J. D.: What's Real About The Business Cycle? *FRB St. Louis – Review* **4** (2005), 435–52.
[6] Tong, H. and Lim, T. S.: Threshold Autoregression, Limit Cycles and Cyclical Data. *Journal of the Royal Statistical Society* **3** (1980), 245-292.

Appendix

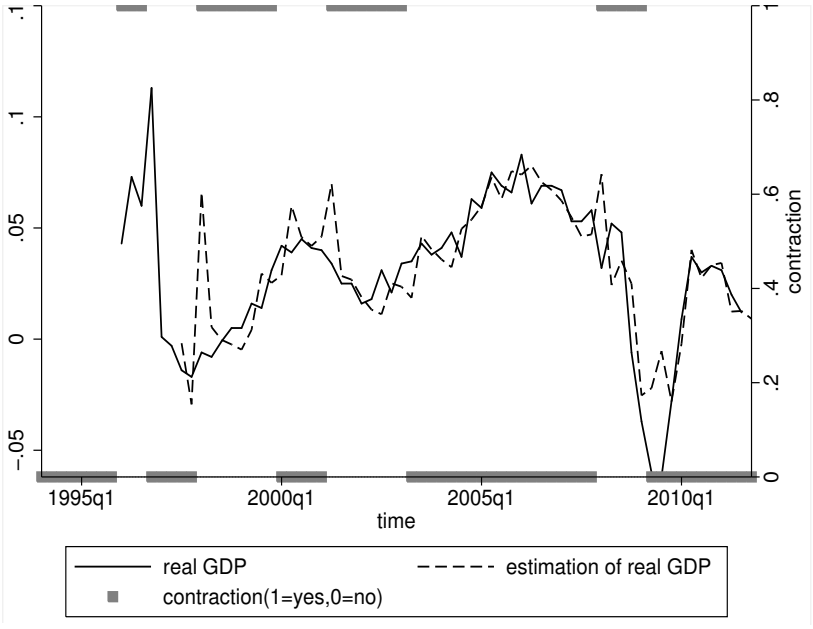


Figure 1 The fit of real GDP

Application of methodology Value at Risk for market risk with normal mixture distribution

Kateřina Zelinková[1]

Abstract. Conditional Value at Risk (CVaR) has been proposed as an alternative, almost coherent, risk measure to Value at Risk (VaR), as it considers expected loss beyond VaR. This paper deals with estimating Value at Risk and conditional Value at Risk under the assumption of mixture normal distribution. We apply mixture normal distribution by assuming that the economy is in various phases of the business cycle. We determinate the both risk measures for market risk, daily returns of popular indices (DAX, CAC, Nikkei and FTSE) over ten years. In the first part, we describe methodology VaR and CVaR and techniques of estimating parameters of probability distributions are presented, i.e. general method of moments and maximum likelihood. Finally, we compare all estimates with each other.

Keywords: mixture distribution, EM algorithm, method of moments, Value at Risk.

JEL Classification: C16, G22, G32
AMS Classification: 91B30

1 Introduction

The Value at Risk (VaR) is defined as the maximum potential loss in value of a portfolio due to adverse market movements, for a given probability. There is also a possibility to refer to VaR as managing risk methodology which is applied widely to modeling credit, operational, market and also insurance risk. Value at Risk is very easy concept; its measurement is a very challenging statistical problem. A good introduction to Value at Risk methodology is provided by the technical document from [11] or by many follow-up books such as [2], [8], [9], [10].

Nevertheless, we can find a lot of VaR criticism, for instance [10]. [3] deals with the features of good risk measure (called coherent) which is defined by four assumption imposed on the ideal risk measure, i.e. monotonicity, sub-additivity, homogeneity and translational invariance. Value at Risk satisfies all these features only in specific case. Specifically, the sub-additivity is violated as far as the portfolio´s profit/loss or portfolio´s return cannot be characterized by some elliptical probability distribution; see [4] for more details. In addition, the VaR says nothing about the loss behind the VaR. Therefore, other risk measures are preferable such as conditional Value at Risk (CVaR) which represents the average of losses exceeding the VaR.

Moreover, if the VaR and CVaR are estimated analytically, the distribution assumption is needed. Normal distribution can be supposed but this assumption results in underestimation of VaR and CVaR due to the existence of fat tails. In facts, to solve this problem, only two approaches seem to be applicable. One can consider Extreme Value Theory (EVT) focused on fitting the tail distribution only which is approximated mostly via general Pareto distribution, the other can apply mixture distribution to fit the empirical distribution the most. The series of studies are devoted the heavy tails [7], [12].

Thus, the aim of paper is estimating Value at Risk and conditional Value at Risk under the assumption of mixture normal distribution and normal distribution. To respect the fat tails, we apply normal mixture distribution. We highlight in this paper that the VaR and CVaR are highly underestimated in that case.

The paper is organized as follows. Section 2 is devoted to the description of normal mixture distribution and methods of its estimate parameters and general Value at Risk methodology under assumption normal mixture distribution. The VaR and CVaR estimates under normal distribution and mixture probability distribution are determined in Section 3 and Section 4 concludes the paper.

[1]VŠB-TU Ostrava, Faculty of Economics, Department of Finance, Sokolská tř. 33, 701 21, katena.zelinkova.@vsb.cz.

2 Value at Risk methodology

In this part, we focus normal mixture distribution and then we describe methods of estimating the parameters of the normal mixture distribution, i.e. maximum likelihood estimation (MLE) and method of moments (MM). Finally, we characterize shortly Value at Risk and Conditional Value at Risk.

2.1 Normal mixture distribution

The financial analysis is based generally on the assumption of normally distributed returns but we know that this assumption is not valid. In most financial markets returns are both skewed and leptokurtic. Hence, a number of alternatives skewed and leptokurtic distributions have been applied. Suitable alternative to normal distribution is a mixture of two or more normal distributions. Normal mixture distribution is described in [6].

We can define the mixture distribution like a probability-weighted sum of other distribution functions. For instance, in a mixture of two normal distributions, there are two regimes for returns: one where the return has mean μ_1 and variance σ_1^2 and another where the return has mean μ_2 and variance σ_2^2. The parameter π is the probability for first regime, so the second regime occurs with probability 1- π. The mixture distribution of just two normal densities is defined by

$$G(x) = \pi F(x:\mu_1,\sigma_1^2) + (1-\pi)F(x:\mu_2,\sigma_2^2), \quad 0 \prec \pi \prec 1, \tag{1}$$

where $F(x:\mu_i,\sigma_i^2)$ denotes the normal distribution function with mean μ_i and variance σ_i^2, for i = 1,2, and where π is the probability associated with the normal component with mean μ_1 and variance σ_1^2. We suppose that the expectation of sample is zero for both distribution, so $\mu_1 = \mu_2 = 0$. In this case the variance of the normal mixture distribution is just the probability- weighted sum of distribution functions, i.e.

$$\sigma^2 = \pi\sigma_1^2 + (1-\pi)\sigma_2^2. \tag{2}$$

The skewness is zero and the kurtosis is given by

$$\kappa = 3\left(\frac{\pi\sigma_1^4 + (1-\pi)\sigma_2^4}{\left[\pi\sigma_1^2 + (1-\pi)\sigma_2^2\right]^2}\right). \tag{3}$$

2.2 Mixture parameter estimation

The estimation of the mixture parameter can be via two methods: maximum likelihood method and method of moments.

Maximum likelihood estimation is a general method for estimating the parameters of a distribution. This method is used extensively because maximum likelihood estimators are consistent. That is, the distribution of the estimator converges to the true value of the parameter as the sample size increases. For estimating the parameter of mixture distribution via method maximum likelihood is used the EM algorithm. The Expectation maximization (EM) algorithm is an efficient iterative procedure to compute the Maximum Likelihood (ML) estimate in the presence of missing or hidden data. Each iteration of the EM algorithm consists of two processes: The E-step, and the M-step. The E-step is the calculation of the expected log likelihood given the current estimates of and given some distribution on the values of the latent variable. In the M-step, the likelihood function is maximized under the assumption that the missing data are known. EM algorithm is described in [1].

As the number of distributions in the mixture increases the probability weight on some of these components can become extremely small. However, in finance it is seldom necessary to use more than two or three components in the mixture, since financial asset return distributions are seldom so irregular as to have multiple modes. In this approach we equate the first few sample moments (one moment for each parameter to be estimated) to the corresponding theoretical moments of the normal mixture distribution. *The method of moments* in general provides estimators which are consistent but not as efficient as the maximum likelihood ones. If we will be apply method of moments for estimate the parameters of a normal mixture distribution we equate to the first four sample moments by changing the parameters of the normal mixture distribution. The vector of probability weights is denoted by $\pi = (\pi_1,.....,\pi_m)$ where $\sum_{i=1}^{m} \pi_i = 1$. The non-central moments are

$$
\begin{aligned}
M_1 &= E[X] = \sum_{i=1}^{m} \pi_i \mu_i \\
M_2 &= E[X^2] = \sum_{i=1}^{m} \pi_i (\sigma_i^2 \mu_i^2) \\
M_3 &= E[X^3] = \sum_{i=1}^{m} \pi_i (3\mu_i \sigma_i^2 + \mu_i^3) \\
M_4 &= E[X^4] = \sum_{i=1}^{m} \pi_i (3\sigma_i^4 + 6\mu_i^2 \sigma_i^2 + \mu_i^4).
\end{aligned} \tag{4}
$$

And the mean (μ), variance (σ^2), skewness (τ) and kurtosis (κ) are

$$
\begin{aligned}
&\mu = E[X] = M_1 \\
&\sigma^2 = E[(X-\mu)^2] = M_2 - M_1^2 \\
&\tau = \sigma^{-3} E[(X-\mu)^3] = \sigma^{-3}(M_3 - 3M_1 M_2 + 2M_1^3) \\
&\kappa = \sigma^{-4} E[(X-\mu)^4] = \sigma^{-4}(M_4 - 4M_1 M_3 + 6M_1^3 M_2 + 2M_1^4).
\end{aligned} \tag{5}
$$

The characteristics of particular moments are described in various books for example [1].

2.3 Value at Risk and Conditional Value at Risk

Value at Risk is defined as the smallest loss for the predicted level of probability for a given time interval. It is a function of two parameters, i.e. the risk horizon and the confidence level. We can also characterize the Value at Risk as a one-sided confidence interval of potential loss of portfolio value for a given holding period, which can be written:

$$
F(x) = P\left(X \le -VaR_{\alpha,\Delta t}(x)\right) = \alpha, \tag{6}
$$

where $F(x)$ is cumulative distribution function, α is significance level and Δt is holding period or risk horizon.

For normal distribution we can write VaR estimation as

$$
VaR = \Phi^{-1}(1-\alpha)\sigma - \mu \tag{7}
$$

where Φ is standard normal distribution function, μ expectation and σ standard deviation. We can determinate VaR from the mixture distribution function thus

$$
G(x) = \pi P\left(X \prec (x_a - \mu_1)\sigma_1^{-1}\right) + (1-\pi)\left(X \prec (x_a - \mu_2)\sigma_2^{-1}\right) = \alpha. \tag{8}
$$

X is normal variable, we know its quartiles. That is, we know everything in the above identity except the mixture quantile x_α. We can find the mixture quantile using an iterative approximation method such as the goal programming and VaR for mixture distribution is $VaR_\alpha = -x_\alpha$.

Conditional Value at Risk informs what the losses would exceed this level. Conditional VaR can be generally defined in the form of

$$
CVaR_{\alpha,\Delta t} = -\alpha^{-1} \int_{-\infty}^{VaR_{\alpha,\Delta t}} x f(x)\, dx, \tag{9}
$$

where $f(x)$ is density function. CVaR is computed by assuming normal distribution as

$$
CVaR_\alpha = \alpha^{-1} \varphi\left(\Phi^{-1}(\alpha)\right)\sigma - \mu \tag{10}
$$

where φ and Φ is standard normal density and distribution function, μ expectation and σ standard deviation. We can write formula for CVaR by assuming the mixture distribution as

$$
CVaR_\alpha = \alpha^{-1} \sum_{i=1}^{n} \left(\pi_i \sigma_i \varphi\left(\sigma_i^{-1} x_\alpha\right)\right) - \sum_{i=1}^{n} \pi_i \mu_i. \tag{11}
$$

3 Determination Value at Risk for mixture distributions

We will apply Value at Risk for four equities indices (CAC 40, DAX, FTSE 100 and NIKKEI 25). We will estimate VaR and CVaR at 99.9% and 99.5% confidence level over one day risk horizon, because the financial institutions determinate capital requirement for market risk for confidence level (banks at 99.9% and insurance companies at 99.5% confidence level.) We prefer one day risk horizon because it is different in financial institutions (for banks is 10 days and insurance companies 1 year horizon). The period sample is between January 2002 and December 2011 and we take the daily log returns for the stock indices over the whole period. The basic numerical characteristics of individual stock returns, especially the mean, standard deviation, kurtosis, skewness are shown in the following table. Histograms of empirical values you can see in Figure 1 – 4.

Variable	Mean	Volatility	Skewness	Kurtosis
CAC 40	-0.0145%	1.5946%	0.0867	5.2427
DAX	0.0044%	1.6478%	0.0678	4.4742
FTSE 100	0.0026%	1.3357%	-0.1195	6.4006
NIKKEI	-0.0076%	1.5748%	-0.4764	7.7677

Table 1 Moments of equities indices

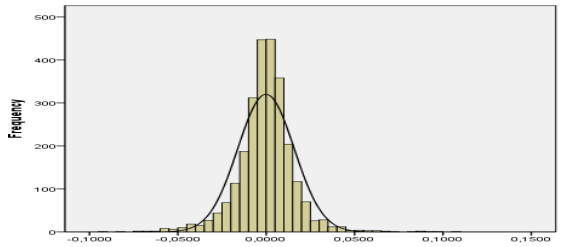


Figure 1 Log returns of CAC

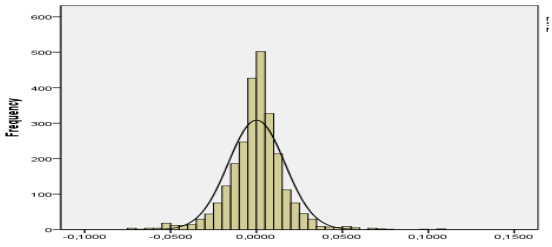


Figure 2 Log returns of DAX

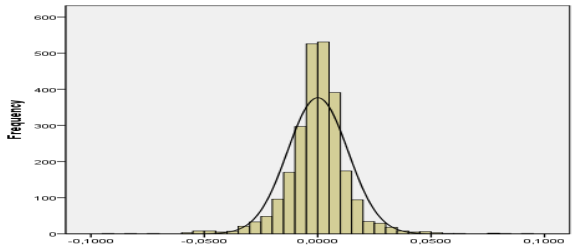


Figure 3 Log returns of FTSE

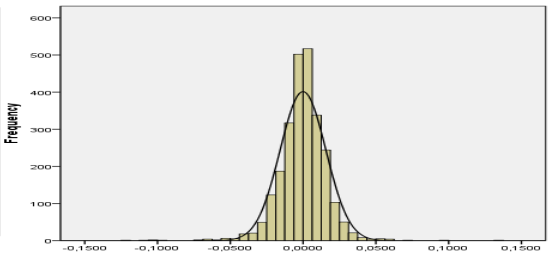


Figure 4 Log returns of NIKKEI

Firstly, we estimate parameters of normal mixture distributions. We apply the method of moment and EM algorithm to fit a mixture of two normal distributions to the daily returns for the four equities indices (CAC 40, DAX, FTSE 100 and NIKKEI 25). We can see estimated parameters in Table 2.

	Method of moments				Maximum likelihood			
	CAC 40	DAX	FTSE 100	NIKKEI	CAC 40	DAX	FTSE 100	NIKKEI
μ_1	0.026%	0.342%	-0.214%	0.007%	-0.045%	0.038%	0.064%	0.081%
μ_2	0.065%	-0.239%	0.082%	0.059%	0.494%	-0.074%	0.052%	0.047%
σ_1	2.085%	1.672%	1.559%	1.020%	1.850%	1.907%	1.559%	1.255%
σ_2	0.849%	0.897%	0.188%	2.256%	0.870%	0.830%	0.588%	0.867%
π_1	0.342	0.592	0.723	0.658	0.650	0.230	0.810	0.380
π_2	0.658	0.408	0.277	0.342	0.350	0.770	0.190	0.620

Table 2 Estimated parameters of normal mixture distribution

We can see differences for various methods of estimate parameters. The highest differences are in parameters π_1. Finally, VaR and CVaR were determined under assumption normal mixture distribution and normal distribution. We use significance levels α = 0.5% and 0.1% and time horizon one day. For each of the four risk factor we use Solver to back out the normal mixture VaR from the equation (8). We calculate VaR model with estimate parameters via method of maximum likelihood and method of moments. The results are in the next tables.

VaR model	Normal Mixture		Normal	Normal Mixture		Normal
	MM	MLE		MM	MLE	
Significance	0.1%			0.5%		
CAC 40	7.76%	8.02%	4.91%	4.58%	4.71%	4.09%
DAX	8.06%	8.32%	5.10%	4.76%	4.64%	4.25%
FTSE 100	6.53%	6.27%	4.13%	3.86%	3.98%	3.44%
NIKKEI	7.68%	7.94%	4.86%	4.53%	4.66%	4.05%

Table 3 The results VaR model

Value VaR under assumption normal distribution is much significantly lower than VaR under assumption normal mixture distribution. The value of VaR is so undervalued, which also leads held by the low economic capital to cover potential risks.

CVaR model	Normal Mixture		Normal	Normal Mixture		Normal
	MM	MLE		MM	MLE	
Significance	0.1%			0.5%		
CAC 40	6.97%	7.22%	5.39%	6.37%	6.49%	5.68%
DAX	7.13%	8.71%	5.55%	6.56%	6.68%	5.85%
FTSE 100	6.08%	6.08%	4.50%	5.32%	5.44%	4.75%
NIKKEI	6.89%	6.89%	5.31%	6.28%	6.40%	5.61%

Table 4 The results CVaR model

The same results we conclude according to the CVaR results. Also in this case, it is obvious that the CVaR estimates are highly underestimated. Thus, the importance of applying normal mixture distribution to quantify the risk measure in the form of VaR or CVaR is obvious and we can highly recommend it.

4 Conclusion

The paper deals with quantification of risk measure using Value at Risk methodology on market risk for four equities indices (CAC 40, DAX, FTSE 100 and NIKKEI 25). Firstly, we characterize normal mixture distribution and determination VaR and CVaR. Subsequently, we estimated parameters mixture distribution via the maximum likelihood method and method of moments. Finally, we determined VaR and CVaR at 99.9% and 99.5% confidence level over one day risk horizon and we compared estimates of both risk measures on the assumption that normal distribution with estimates under assumption normal mixture distribution.

We know that asset returns tend to be skewed and heavy tails. Only just normal mixture distribution can model heavy tails and we supposed the fat tail of probability distribution and therefore we applied normal mixture distribution. Thus, we do not take fat tails into account it can lead to the very imprecise and very different Value at Risk estimates resulting in insufficient capital which should cover the loss.

Acknowledgements

This paper was solved within the project SP2012/19 Estimating rating models and analysis of the influence qualitative factors on the rating evaluation.

References

[1] Alexander, C.: *Quantitative methods in finance*. John Wiley& Sons Inc , Chichester, 2008
[2] Alexander, C.: *Value at Risk models*. John Wiley& Sons Inc , Chichester, 2008.

[3] Artzner, P., Delbaen, F., Eber, J.–M., and Heath, D. *Coherent Measures of Risk. Mathematical Finance* **9**, (1999), 203–228.
[4] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. *Statistics of Extremes: Theory and Applications*. John Wiley& Sons Inc, Chichester, 2004.
[5] Danielsson, J. *The Value at Risk: Key Issues in the Implementation of Market Risk.* Risk Books, London, 2007.
[6] Frühwirth-Schnatter S. *Finite Mixture and Markov Switching Models.*Springer, New York, 2006.
[7] Harmantzis F.C., Miao L. and Y. Chien. *Empirical study of value at risk and expected shortfall model with heavy tails. Journal of Risk Finance* **7**, (2006), 117-135.
[8] Hull, J. C. *Risk management and Financial Institutions*. Pearson Education, New Jersey, 2007.
[9] Jorion, P. *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill, New York, 2007.
[10] Knobloch, A.P. *Value at Risk: Regulatory and Other Applications, Methods, and Criticism. Risk management*. Springer Verlag, Berlin, 2005, 99-124.
[11] Morgan, J.P. RiskMetricsTM Technical Document, New York, 1996.
[12] Neftci, S. *Value at Risk Calculations, Extreme Events, and Tail Estimation. The Journal of Derivates* **3**, (2000), 23-27.

Modelling the sequential real options under uncertainty and vagueness (fuzzy-stochastic approach)

Zdeněk Zmeškal[1]

Abstract
The basic intention of the paper is to propose and verify the sequential real options model under vagueness conditions. The sequential real options are specific real option type. Sequential options are special type of generalised switch option. The sequential problems could be decomposed and solved as several subsequent stages. And subsequent value is underlying asset of the computed stage. Therefore, the sequential option model is compound options on the stages values. The input data uncertainty and vagueness in a form of fuzzy-stochastic distribution function is considered. Sequential fuzzy-stochastic model is proposed. Illustrative example is presented.

Keywords: Real options; Switch option; Discrete Binomial Model; Stochastic Dynamic Programming; Sensitivity analysis; Fuzzy-stochastic model; Fuzzy number

JEL Classification: C6, C 44, C53, F2, F21, G1, G11, G15, G2, G21
AMS Classification: 91B25, 91G20, 91G50, 91G60, 91G80

1 Introduction

The real options methodology could be considered to be a generalised approach encompassing risk and flexibility aspects simultaneously in a valuation. We can present publications concerning real options, see e. g. [9], [17], [20], [27], [28], [29], [1], [5], [6], [7], [8], [10], [13]. Relatively new topic of option valuation are fuzzy-stochastic models, examples of papers are e.g. in [25], [26], [22], [23], [2], [3], [4], [19], [21], [15], [16], [17], [24], [28], [29], [30], [31], [32], [33], [14], [17], [11].

Basic approach of valuation under complete market is the replication strategy and no-arbitrage principle using the risk-neutral probability and general principle is the martingale approach. Owing to the economic assets types, random processes complexity and decisions variables and functions, the real options are largely of the American options type, discrete binomial (multinomial) models, with multinomial options to switch. Sequential options are a special type of switching options modelling being a generalised approach of real options modelling. Sequential real options are irreversible and state (phase) options applied especially in the investment decision process, construction; see e.g. [20], [30], [31], [32]. The investment (switching) cost is influential input parameter of sequential option valuation. A sensitivity of the value and results are the important aspect of the decision-making and analysis.

The intention of the paper is to analyse the possibilities of the sensitivity of the sequential options value on the investment cost. The fuzzy-stochastic approach is investigated. Methodology and notation is derived and linked to the papers of [26], [31], [32].

2 Valuation procedure of an American sequential option with switching cost

The sequential problems could be decomposed and solved as several subsequent stages. And subsequent value is underlying asset of the computed stage, see Fig. 1.

[1] VSB – Technical University of Ostrava, Faculty of Economics, Department of Finance, Sokolská 33, Ostrava, Zdenek.Zmeskal@vsb.cz.

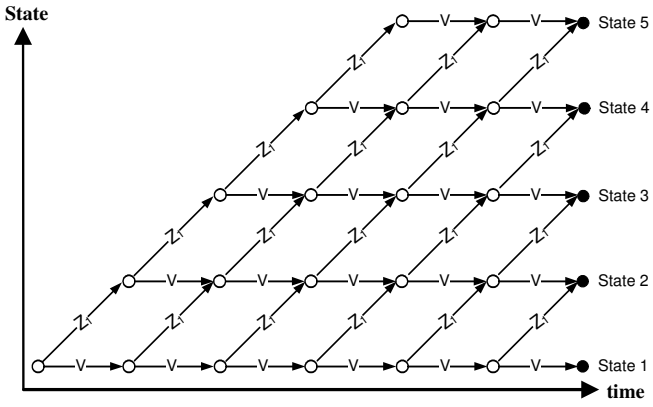


Figure 1 Sequential real option presentation

Therefore, problem is in detail a compound option on the stages values. In every stage there are only two decision possibilities: go to subsequent stage *m-1* or to stay in stage *m*. Applying notation in [31] the sequential options valuations equations are following.

$$V_N^m = \max\left[x_0^{m-1} - C_{m,m-1} + \beta \cdot \widehat{E}\left(V_{N-1}^{m-1}\right), x_0^m + \beta \cdot \widehat{E}\left(V_{N-1}^m\right)\right], \tag{1}$$

$$V_{N-k}^m = \max\left[x_k^{m-1} - C_{m,m-1} + \beta \cdot \widehat{E}\left(V_{N-1-k}^{m-1}\right), x_k^m + \beta \cdot \widehat{E}\left(V_{N-1-k}^m\right)\right], \tag{2}$$

$$V_1^m = \max\left[x_{N-1}^m + \beta \cdot V_0^m\right] \tag{3}$$

Here V_{N-k}^m is value for *N-k* periods to final period and x_k^{m-1} is cash flow in the particular period k and stage *m-1*, $C_{m,m-1}$ are investment (switch) cost from stage *m* to *m-1*, $\beta_t = (1+R)^{-t}$ is discount factor, $\widehat{E}\left(V_{N-1-k}^q\right)$ is risk-neutral expected value of subsequent stage and *N-k-1*is time to final stage.

2.1 Sequential option valuation procedure description

Valuation procedure of sequential options with investment cost reflecting the stochastic dynamic programming on the Bellman's principle expressed by recurrent equations, under the discrete binomial model and risk-neutral probability and geometric Brown motion is possible to describe in following steps.

a) The determination of the risk-neutral growth parameter $\widehat{g}$.

b) Cash flow modelling so as an underlying asset. Subjective approach by virtue of expert estimation and forecast. Objective approach on the basis of statistical estimation and forecasting of random process. In the case of Brown's geometrical process due to Cox, Ross, Rubinstein (1979) calibration,
$x_{t+1,s+u}^u = x_{t,s} \cdot U$; $x_{t+1,s+d}^d = x_t \cdot D$. Here $U = e^{\sigma\cdot\sqrt{dt}}$, $D = e^{-\sigma\cdot\sqrt{dt}}$.

c) At the beginning of the last phase the value for the last phase is $V_{1,s}^q$, here *s* is state and *q* is mode.

d) By backward recurrent procedure from the end of binomial tree to the beginning for states *s* and modes *q* of particular period in accordance with the generalised recurrent Bellman's stochastic equations (1), (2), (3) correspondingly values are calculated. Here $\widehat{p}$ is risk-neutral probability of up movement and $\widehat{q} = 1 - \widehat{p}$ is risk-neutral probability of down movement. Valuation formula for the last phase,
$$V_1^m = \max\left[x_{N-1}^m + \beta \cdot V_0^m\right].$$
Valuation formula for other periods by virtue of the recurrent procedure,
$$V_{N-k}^m = \max\left[x_k^{m-1} - C_{m,m-1} + \beta \cdot \widehat{E}\left(V_{N-1-k}^{m-1}\right), x_k^m + \beta \cdot \widehat{E}\left(V_{N-1-k}^m\right)\right].$$
Valuation formula at the beginning of the whole first phase (the first period),
$$V_N^m = \max\left[x_0^{m-1} - C_{m,m-1} + \beta \cdot \widehat{E}\left(V_{N-1}^{m-1}\right), x_0^m + \beta \cdot \widehat{E}\left(V_{N-1}^m\right)\right].$$

e) Identification of the decision variant for state s and time t, $Q_{t,s}$,

$$Q_{t,s} = \arg\max_{q \in S}\left(x_k^{m-1} - C_{m,m-1} + \beta \cdot \hat{E}\left(V_{N-1-k}^{m-1}\right); x_k^m + \beta \cdot \hat{E}\left(V_{N-1-k}^m\right)\right).$$

f) The implementation of the sensitivity analysis concerning the input data.

3 Fuzzy-stochastic elements

For application of fuzzy-stochastic methodology the crucial terms are fuzzy number, fuzzy operations and decomposition principle. A fuzzy set meeting preconditions of normality, convexity, continuity with the upper semi-continuous membership function and closeness and being depicted as the quadruple $\tilde{s} = \left(s^L, s^U, s^\alpha, s^\beta\right)$, where $\phi(\mathrm{x})$ is a non-decreasing function and $\psi(\mathrm{x})$ is a non-increasing function, is called the *T-number*. Let us denote the set of T-numbers on n-dimensional Euclidean space E by $F_T(E)$. *T-number* is defined as follows,

$$\tilde{s} \equiv \mu_{\tilde{s}}(x) = \left\{ \begin{array}{llll} 0 & for & x \le \mathrm{s}^\mathrm{L} - s^\alpha; & \phi(\mathrm{x}) \ for \quad \mathrm{s}^\mathrm{L} - s^\alpha < \mathrm{x} < \mathrm{s}^L; \\ 1 & for & \mathrm{s}^\mathrm{L} \le x \le s^U; & \psi(\mathrm{x}) \ for \quad \mathrm{s}^\mathrm{U} < \mathrm{x} < \mathrm{s}^\mathrm{U} + s^\beta; \\ 0 & for & x \ge \mathrm{s}^\mathrm{U} + s^\beta & \end{array} \right\}.$$

The ε-cut of the fuzzy set $\tilde{s}$, depicted $\tilde{s}^\varepsilon$, is defined as follows. $\tilde{s}^\varepsilon = \left\{\mathrm{x} \in \mathrm{E}; \mu_{\tilde{s}}(x) \ge \varepsilon\right\} = [{}^-s^\varepsilon, {}^+s^\varepsilon]$ where ${}^-s^\varepsilon = \inf\left\{x \in E; \mu_{\tilde{s}}(x) \ge \varepsilon\right\}$, ${}^+s^\varepsilon = \sup\left\{x \in E; \mu_{\tilde{s}}(x) \ge \varepsilon\right\}$.

Application of the Decomposition principle for a function of fuzzy numbers allows expressing the selected fuzzy operations $\tilde{*}$ among fuzzy numbers directly, as follows: $\tilde{w} = \tilde{s} \,\tilde{*}\, \tilde{r} = \bigcup_\varepsilon \varepsilon\left(w^\varepsilon\right) = \bigcup_\varepsilon \varepsilon\left(s^\varepsilon * r^\varepsilon\right)$.

Fuzzy addition $s^\varepsilon + r^\varepsilon = \left[{}^-s^\varepsilon + {}^-r^\varepsilon; {}^+s^\varepsilon + {}^+r^\varepsilon\right]$.

Fuzzy subtract $s^\varepsilon - r^\varepsilon = \left[{}^-s^\varepsilon - {}^+r^\varepsilon; {}^+s^\varepsilon - {}^-r^\varepsilon\right]$.

Fuzzy scalar product $k \cdot s^\varepsilon = \left[k \cdot {}^-s^\varepsilon; k \cdot {}^+s^\varepsilon\right]$ for $k \ge 0$, $k \cdot s^\varepsilon = \left[k \cdot {}^+s^\varepsilon; k \cdot {}^-s^\varepsilon\right]$ for $k < 0$.

Fuzzy multiplication $s^\varepsilon \cdot r^\varepsilon = \left[{}^-s^\varepsilon \cdot {}^-r^\varepsilon; {}^+s^\varepsilon \cdot {}^+r^\varepsilon\right]$ for $\tilde{s} > 0,\ \tilde{r} > 0$,

$s^\varepsilon \cdot r^\varepsilon = \left[{}^-s^\varepsilon \cdot {}^+r^\varepsilon; {}^+s^\varepsilon \cdot {}^-r^\varepsilon\right]$ for $\tilde{s} < 0,\ \tilde{r} > 0$, $s^\varepsilon \cdot r^\varepsilon = \left[{}^+s^\varepsilon \cdot {}^+r^\varepsilon; {}^-s^\varepsilon \cdot {}^-r^\varepsilon\right]$ for $\tilde{s} < 0,\ \tilde{r} < 0$.

Fuzzy division $s^\varepsilon : r^\varepsilon = \left[{}^-s^\varepsilon : {}^+r^\varepsilon; {}^+s^\varepsilon : {}^-r^\varepsilon\right]$ for $\tilde{s} > 0,\ \tilde{r} > 0$, $s^\varepsilon : r^\varepsilon = \left[{}^+s^\varepsilon : {}^+r^\varepsilon; {}^-s^\varepsilon : {}^-r^\varepsilon\right]$ for $\tilde{s} < 0,\ \tilde{r} > 0$,

$s^\varepsilon : r^\varepsilon = \left[{}^+s^\varepsilon : {}^-r^\varepsilon; {}^-s^\varepsilon : {}^+r^\varepsilon\right]$ for $\tilde{s} < 0,\ \tilde{r} < 0$.

Fuzzy max, $\max\left(s^\varepsilon\right) = \left[\max {}^-s^\varepsilon; \max {}^+s^\varepsilon\right]$.

Here $\tilde{s} > 0$ is positive fuzzy number, positive $\tilde{s}: \{x; \text{for which } \mu_{\tilde{s}} \ge 0\}$ and simultaneously $x \in E^+$ (set of positive numbers), negative $\tilde{s}: \{x; \text{for which } \mu_{\tilde{s}} \ge 0\}$ and simultaneously $x \in E^-$ (set of negative numbers). *Decomposition principle* (Resolution identity) is defined as follows, $\mu_{\tilde{s}}(y) = \sup_\varepsilon\left\{\varepsilon \cdot I_{\tilde{s}^\varepsilon}; y \in \tilde{s}^\varepsilon\right\}$ for any $y \in E$ and $\varepsilon \in [0;1]$, where $\tilde{s}^\varepsilon = [{}^-s^\varepsilon, {}^+s^\varepsilon]$ is ε-cut, $\mathrm{y} = \mathrm{f(x)}$, ${}^-s^\varepsilon(\mathrm{x}) = \min_{x \in \tilde{x}^\varepsilon \subset E} \mathrm{f(x)}$, ${}^+s^\varepsilon(\mathrm{x}) = \max_{x \in \tilde{x}^\varepsilon \subset E^n} \mathrm{f(x)}$.

Here $I_{\tilde{s}^\varepsilon}$ is the characterisation function, $I_{\tilde{s}^\varepsilon} = \left\{1 \ if \ \mathrm{y} \in [{}^-s^\varepsilon, {}^+s^\varepsilon]; 0 \ if \ \mathrm{y} \notin [{}^-s^\varepsilon, {}^+s^\varepsilon]\right\}$.

4 Fuzzy-stochastic sequential real option model

There is several variants of the fuzzy input data: partial fuzzy input data (investment cost, volatility (up index, down index),, underlying asset, risk free –rate), full fuzzy input data (every parameters are in fuzzy values given). We can generally express, applying the decomposition principle, ε-cut of fuzzy number $\tilde{f}_t$, composed from two fuzzy numbers $\tilde{G}_t$ and $\tilde{H}_t$ by relation fuzzy maximization, as follows,

$$\left[f_t^\varepsilon\right] = \max\left[G_t^\varepsilon; H_t^\varepsilon\right] = \max\left[\left({}^-G_t^\varepsilon; {}^-H_t^\varepsilon\right), \left({}^+G_t^\varepsilon; {}^+H_t^\varepsilon\right)\right] = \left[\max\left({}^-G_t^\varepsilon; {}^-H_t^\varepsilon\right), \max\left({}^+G_t^\varepsilon; {}^+H_t^\varepsilon\right)\right] = \left[{}^-f_t^\varepsilon; {}^+f_t^\varepsilon\right].$$

Here ${}^-f_t^\varepsilon = \max\left({}^-G_t^\varepsilon; {}^-H_t^\varepsilon\right)$, ${}^+f_t^\varepsilon = \max\left({}^+G_t^\varepsilon; {}^+H_t^\varepsilon\right)$. And ${}^-G_t^\varepsilon = \min G_t^\varepsilon$, ${}^-H_t^\varepsilon = \min H_t^\varepsilon$, furthermore ${}^+G_t^\varepsilon = \max G_t^\varepsilon$, ${}^+H_t^\varepsilon = \max H_t^\varepsilon$.

Applying previous common fuzzy maximization relation with fuzzy investment (switch) cost, we can formulate following recurrent formulas,

$$\left(V_N^m\right)^{\varepsilon} = \left[{}^{-}\left(V_N^m\right)^{\varepsilon}; {}^{+}\left(V_N^m\right)^{\varepsilon} \right], \tag{4}$$

where ${}^{-}\left(V_N^m\right)^{\varepsilon} = \max\left[x_0^{m-1} - {}^{+}C_{m,m-1}^{\varepsilon} + \beta \cdot \widehat{E}\left(V_{N-1}^{m-1}\right); x_0^m + \beta \cdot \widehat{E}\left(V_{N-1}^m\right)\right]$ and

$${}^{+}\left(V_N^m\right)^{\varepsilon} = \max\left[x_0^{m-1} - {}^{-}C_{m,m-1}^{\varepsilon} + \beta \cdot \widehat{E}\left(V_{N-1}^{m-1}\right); x_0^m + \beta \cdot \widehat{E}\left(V_{N-1}^m\right)\right];$$

$$\left(V_{N-k}^m\right)^{\varepsilon} = \left[{}^{-}\left(V_{N-k}^m\right)^{\varepsilon}; {}^{+}\left(V_{N-k}^m\right)^{\varepsilon} \right], \tag{5}$$

where ${}^{-}\left(V_{N-k}^m\right)^{\varepsilon} = \left[x_k^{m-1} - {}^{+}C_{m,m-1}^{\varepsilon} + \beta \cdot \widehat{E}\left(V_{N-1-k}^{m-1}\right); x_k^m + \beta \cdot \widehat{E}\left(V_{N-1-k}^m\right)\right]$ and

$${}^{+}\left(V_{N-k}^m\right)^{\varepsilon} = \max\left[x_k^{m-1} - {}^{-}C_{m,m-1}^{\varepsilon} + \beta \cdot \widehat{E}\left(V_{N-1-k}^{m-1}\right); x_k^m + \beta \cdot \widehat{E}\left(V_{N-1-k}^m\right)\right];$$

$$\left(V_1^m\right)^{\varepsilon} = \left[{}^{-}\left(V_1^m\right)^{\varepsilon}; {}^{+}\left(V_1^m\right)^{\varepsilon} \right] = \left[\left(x_{N-1}^m + \beta \cdot V_0^m\right); \left(x_{N-1}^m + \beta \cdot V_0^m\right)\right] \tag{6}$$

5 Illustrative example of fuzzy-stochastic sequential real option model with fuzzy investment cost

The sequential fuzzy-stochastic real option value with fuzzy investment cost will be calculated. The binomial model, American option; replication value strategy, risk-neutral approach; expected present value objective function will be employed. The applied model is of three-phase type. We suppose that random cash flow (underlying asset) follows geometric Brown process. The model is base on the equation (4), (5) and (6).

Input data of the model are following: risk-free rate r = 10%; up-movement index U = 1,2. The price of underlying cash flow for last phase is 10 c.u., for the second phase 4 c.u. and the first phase 1 c.u. Risk-neutral probability of up movement is $\hat{p}$ = 72,73 % and down movement is $\hat{q} = 1 - \hat{p} = 27{,}27\%$. Fuzzy investment cost from the first to the second phase is $\tilde{s}^{12} = \left(s^L, s^U, s^{\alpha}, s^{\beta}\right) = (5, 5, 1, 1)$, from the second phase to the third phase is $\tilde{s}^{23} = \left(s^L, s^U, s^{\alpha}, s^{\beta}\right) = (6, 6, 2, 2)$.

The computes results are in Tab. 1. We can see sensitive value of sequential option and optimal starting stage depending on fuzzy epsilon cut.

epsilon	Value		Starting phase	
	min	max	min	max
1	4,073	4,073	V2	V3
0,75	3,515	4,678	V2	V3
0,5	2,995	5,393	V2	V3
0,25	2,52	6,116	V2	V3
0	2,044	6,844	V2	V3

Tab. 1 Fuzzy value of sequential real option in epsilon cut

6 Conclusions

The fuzzy-stochastic sequential option was proposed and investigated in the paper. The fuzzy-stochastic sequential real option model was formulated. Special partial fuzzy-stochastic model for fuzzy investment cost was introduced and illustrative example was presented. The model should be considered to be a generalised sensitivity sequential real option valuation.

Acknowledgements

This paper has been elaborated in the framework of the IT4Innovations Centre of Excellence project, reg. no. CZ.1.05/1.1.00/02.0070 supported by Operational Programme 'Research and Development for Innovations' funded by Structural Funds of the European Union and state budget of the Czech Republic.

References

[1] Brandao, L. E., Dyer, J. S.: Decision Analysis and Real Options: A Discrete Time Approach to Real Option Valuation. *Annals of Operations Research* **135** (2005), p. 21.

[2] Carlsson, C., Fuller, R.: A fuzzy approach to real option valuation. *Fuzzy Sets and Systems* **139** (2003), pp. 297-312.

[3] Cheng Few-Lee, Tzeng Gwo-Hsiung, Wang Shin-Yun,: A Fuzzy Set Approach for Generalized CRR Model: An Empirical Analysis of S&P 500 Index Options. *Review of Quantitative Finance and Accounting*, **25** (2005) , pp. 255–275.

[4] Chrysafis, K. A., Papadopoulos, B. K. : On theoretical pricing of options with fuzzy estimators. *Journal of Computational and Applied Mathematics.* **223/ 2**(2009) , pp. 552-566

[5] Čulík, M.: Valuing of real options portfolio and its influence on project value. In: *22nd International Conference on Mathematical Methods in Economics*. 2004, pp. 56-62.

[6] Čulík, M.: Real option application for modular project valuation. In: *24th International Conference on Mathematical Methods in Economics*. 2006, pp. 123-130.

[7] Čulík, M.: Emissions trading: investment decision-making and real option analysis (simulation approach). *4th International Scientific Conference on Managing and Modelling of Financial Risk*, Ostrava. 2008, pp 11-18

[8] Čulík, M.: Flexibility and project value: interactions and multiple real options. In. 3rd *Global Conference on Power Control and Optimization*, Gold Coast, AUSTRALIA. POWER CONTROL AND OPTIMIZATION Book Series: AIP Conference Proceedings . 2010 Volume: 1239, pp 326-334

[9] Dixit, A. K., Pindyck, R.S.: *Investment under Uncertainty*. Princeton University Press, 1994.

[10] Dluhošová, D.: An analysis of financial performance using the EVA method. *Finance a úvěr- Czech Journal of Economics and Finance* **11-12** (2004).

[11] Figa-Talamanca, G. M., Guerra, M. L., Stefanini, L.: Market Application of the Fuzzy-Stochastic Approach in the Heston Option Pricing Model. *Finance a Úvěr - Czech Journal of Economics and Finance* **62,** Issue 2, (2012), pp. 162-179

[12] Guerra, M. L., Sorini, L., Stefanini, L.: Parametrized fuzzy numbers for option pricing. *IEEE International Conference on Fuzzy Systems*, London, 2007, VOLS 1-4, pp 727-732.

[13] Guthrie, G.: *Real Options in Theory and Practice*. Oxford University Press, 2009.

[14] Liyan Han, Wenli Chen: The Generalization of λ-Fuzzy Measures with Application to the Fuzzy Option. In: Wang, L. et al. (Eds.) Proceedings: *Fuzzy Systems and Knowledge Discovery*, 2006, LNAI 4223, pp. 762–765, Springer-Verlag Berlin Heidelberg.

[15] Muzzioli, S., Reynaerts, H.: The solution of fuzzy linear systems by non-linear programming: a financial application. *European Journal of Operational Research* **177** (2007), pp 1218–1231

[16] Muzzioli, S., Torricelli, C.: A multi-period binomial model for pricing options in a vague world. *Journal of Economic Dynamics and Control* **28** (2004), pp. 861-887.

[17] Smith, J. E., Nau, R. F.: Valuing risky projects: Option pricing theory and decision analysis. *Management Science* **14** (1995), p. 795.

[18] Thiagarajaha, K., Appadoob, S.S., Thavaneswaranc, A.: Option valuation model with adaptive fuzzy numbers, *Computers and Mathematics with Applications* **53** (2007), pp. 831–841.

[19] Thavaneswaran, A., Appadoo, S.S., Frank, J.: Binary option pricing using fuzzy numbers. *Applied Mathematics Letters*, *In Press, Corrected Proof, Available online 13 April (2012)*

[20] Trigeorgis, L.: *Real Options - Managerial Flexibility and Strategy in Resource Allocation*. Harvard University, 1998.

[21] Wu, H.-C.: Using fuzzy sets theory and Black–Scholes formula to generate pricing boundaries of European options. *Applied Mathematics and Computation*, **185/ 1** (2007) , pp. 136-146.

[22] Yoshida, Y.: A discrete-time model of American put option in an uncertain environment. *European Journal of Operational Research* **151** (2002), pp 153-166.

[23] Yoshida, Y.: The valuation of European options in uncertain environment. *European Journal of Operational Research* **145** (2003), pp. 221-229

[24] Zhang Jinliang, Du Huibin, Tang Wansheng: *Pricing R&D option with combining randomness and fuzziness. Computational intelligence*, PT 2, Proceedings Book Series: Lecture notes in artificial intelligence, 2006, Volume: 4114, pp 798-808

[25] Zmeškal, Z.: Fuzzy-stochastic estimation of a firm value as a call option. *Finance a úvěr - Czech Journal of Economics and Finance* **3** (1999), pp. 168-175 .

[26] Zmeškal, Z.: Application of the fuzzy - stochastic Methodology to Appraising the Firm Value as a European Call Option. *European Journal of Operational Research* **135/2** (2001), pp. 303-310.

[27] Zmeškal, Z.: Hedging strategies and financial risks. *Finance a Úvěr - Czech Journal of Economics and Finance* **54** (2004), pp. 50-63.

[28] Zmeškal, Z.: Approach to Real Option Model Application on Soft Binomial Basis Fuzzy - stochastic approach. In: *23rd International Conference on Mathematical Methods in Economics*. 2005, pp. 433-439.
[29] Zmeškal, Z.: Real option applications based on the generalised multinomial flexible switch options methodology. In: *24th International Conference on Mathematical Methods in Economics*. 2006, pp. 545-553.
[30] Zmeškal, Z.: Application of the American Real Flexible Switch Options Methodology A Generalized Approach. *Finance a Úvěr - Czech Journal of Economics and Finance* **58** (2008), pp. 261-275.
[31] Zmeškal, Z.: Generalised soft binomial American real option pricing model (fuzzy–stochastic approach). *European Journal of Operational Research*. **207**/ **2** (2010) , pp 1096-1103**.** doi:10.1016/j.ejor.2010.05.045.
[32] Zmeškal, Z.: Real compound sequential and learning options modelling. In: *28th International Conference on Mathematical Methods in Economics*. 2010, pp. 688-693.
[33] Zmeškal, Z.: Investigation of the Real Switch Option Value Sensitivity. In: *29th International Conference on Mathematical Methods in Economics*. 2011, pp. 774-779.

Are fast food chains really that efficient? A case study on crew optimization.

Jan Zouhar[1], Irena Havlová[2]

Abstract.

Fast food chains have long been setting an example in operations management efficiency; however, as we show in the paper, there is still room for improvement. Our case study deals with crew optimization in a particular fast food restaurant. First, we describe the current practice in crew scheduling—which is mostly done manually by the employees. Next, we provide two alternative MILP formulations that can be used to find the optimal schedule under the given criteria; these formulations take into account the specific crew structure of the fast food restaurant (full-time vs. part-time employees, underage temporary jobs etc.), as well as its specific criteria. We discuss and compare the computational tractability of the proposed models for small-scale applications and comment on solution concepts for large-scale ones. Finally, we evaluate the restaurant's crew scheduling efficiency by comparing the quality of the actual (empiric) schedules to those obtained by MILP.

Keywords: supply chain efficiency, crew scheduling, mixed integer linear programming.

JEL classification: M54, L66
AMS classification: 90B70

1 Introduction

The huge success of fast food giants (such as McDonald's, Burger King or Kentucky Fried Chicken) has usually been attributed to their exceptional operations efficiency. Indeed, fast food chains were among the first companies outside the manufacturing industry that utilized modern production management principles, such as lean production and Just-in-Time (JIT) strategy. Undoubtedly, the sheer idea of a fast food restaurant logically encompasses some of JIT practices, e.g. short lead times, quick setups and small-lot production—but many fast food chains did not stop there. For instance, a couple years ago, McDonald's started using Kanban cards in order to establish a pull production system, and developed a new type of bun toaster to cut down on production time.

Another source of fast food restaurants' cost efficiency is the high degree of work standardization. Among other things, it allows them to hire staff with minimal training, making temporary jobs at the restaurant suitable for underage workers (especially students), who have few alternative job opportunities. As noted in [3], fast food industry exhibits a very high proportion of minimum-wage workers. Moreover, the presence of many part-time employees and/or temporary jobs increases the flexibility of personnel schedules; in other words, it creates the opportunity to adjust the amount of manpower at a particular time to predicted staff requirements (which reduces potential costs of staff shortage/overage).

However, as we demonstrate using a case study of a Czech fast food restaurant[3], these potential benefits from schedule flexibility are not exploited efficiently by the fast food restaurants' executives. The main reason is that, to our knowledge, crew scheduling is mostly done manually[4]—which both takes

[1]Univ. of Economics, Prague, Dept. of Econometrics, nám. W. Churchilla 4, 130 67 Praha 3, e-mail: zouharj@vse.cz
[2]Univ. of Economics, Prague, Dept. of Econometrics, nám. W. Churchilla 4, 130 67 Praha 3, e-mail: xhavil0@vse.cz
[3]Neither the name of the fast food chain nor the location of the restaurant can be disclosed due to confidentiality reasons.
[4]This was the case with the particular fast food restaurant we studied; the restaurant's executive reported that it was a common practice within the whole fast food chain. Obviously, there are exceptions to this rule, though probably not in the Czech Republic; [7] reports on a U.S. example of a fast food restaurant that uses an automated crew scheduling system.

a long time even for an experienced schedule manager and results in a sub-optimal schedule, since with more than only a handful of employees it is virtually impossible to find the optimal schedule by hand.

In the rest of the paper, we aim to support this claim using the data from our case study. In §2, we briefly review existing optimization approaches to crew scheduling. In §3, we present a thorough description of the crew scheduling problem we dealt with in our case study. As our case study was of a relatively small scale—the restaurant had 40 employees (full-time, part-time and temporary)—we could use mixed integer linear programming (MILP) as the optimization vehicle. In §4, we propose two quite different MILP formulations of the scheduling problem. Both have their pros and cons: one is easier to implement, the other is more computationally efficient; we provide a brief discussion of both models' computational tractability at the end of §4. Finally, in §5, we evaulate the restaurant's crew scheduling efficiency by comparing the quality of the actual (empiric) schedules to those obtained by MILP.

2 Optimization approaches to crew scheduling

In general, crew scheduling is the process of creating work timetables for a firm's staff that (1) are feasible/satisfactory for individual employees and (2) efficiently cover the demand for manpower over the given time horizon. In all but the simplest cases, finding a good schedule is a very demanding task, and many automated decision-support systems have been developed for different application areas, which mostly come from the transportation, health care, and services industries; an extensive review is given in [5]. Solutions are typicaly obtained using one of the following techniques:

- *Mathematical programming.* Most crew scheduling problems can be described as a MILP problem; however, for large-scale problems and/or problems with complicated constraint structures, MILP models are not computationally tractable. Nevertheless, this is the approach we took in our case study, which is a rather small-scale one. As we show in §4, our solution admits some up-scaling, but for a large-scale version of the problem, heuristics have to be used instead.
- *Constraint programming* (CP). In general, CP is not a very efficient optimization technique; however, as noted in [5], CP is particularly useful for crew scheduling if the problem is highly constrained and/or when any feasible solution will suffice even if it is not optimal.
- *Metaheuristics.* Most metaheuristic approaches can be adapted to crew scheduling problems, e.g. genetic algorithms (see [8]) or simulated annealing (see [1]).
- *Other heuristics.* Apart from the applications of general metaheuristic approaches, a considerable effort has been put into designing tailor-made heuristics for specific types of staff scheduling problems (see e.g. [2]); several of these heuristics contain applications of AI techniques, as in [6].

3 Problem statement

Specifics of crew scheduling in fast food chains. Fast food crew schedules exhibits several features that make it difficult to optimize. Perhaps the most prominent one is crew heterogeneity. As mentioned above, fast food chains rely heavily on part-time and/or temporary employees, some of which are typically underage students. There are several implications of this. Firstly, there are multiple skill levels that have to be taken into account and scheduled in parallel; in our case study, there has to be at least one full-time manager present at each time to supervise the current shift. Secondly, there are no standard shifts, such as regular 8-hour blocks (which is the case in some other industries, and facilitates scheduling to a great extent). Different types of employees allow for and/or require different shift lengths. Finally, part-time employees can specify their *available times*, i.e. times when they can actually be on duty. These times are typically specified only a short time in advance: in our case study, non-full-time workers announce their available times only 14 days ahead. This effectively rules out long-term planning, and (more importantly) repetitive schedule plans; in other words, the schedule has to be re-optimized each week.

Scope of optimization. In our case study, we omit *night shifts* and *managers* from the model. According to the restaurant's executive, both of these are scheduled separately, on a monthly basis (unlike the remaining part of the schedule, which is prepared weekly for the reasons stated above).

Object types and sets. In our statement of the crew scheduling problem, we use the following objects: employees, days, time periods and shifts. These objects, together with the related sets and their notation,

are described in the following list:

E = the set of all **e**mployees. In the fast food chains, only a minor part of the staff is typically made up by *full-time employees*. The rest can be divided into *part-timers*, *temporary jobs*, and *minors* (underage temporary-job employees, mostly students). In the mathematical notation, we can write $E = E_{\text{full}} \cup E_{\text{part}} \cup E_{\text{temp}} \cup E_{\text{min}}$.

D = the set of (numbered) working **d**ays (typically, Sunday through Saturday, i.e. $D = \{1, \ldots, 7\}$).

T = the set of (numbered) **t**ime periods within a single day. (In our case study, these are half-hour periods from 6:00 to 22:00 each day, and therefore $T = \{1, \ldots, 32\}$.)

S = the set of acceptable shifts. In brief, shifts are collections of consecutive time periods; we will describe acceptable shifts in more detail later on.

Scheduling constraints. A feasible schedule must follow be the following conditions:

(*i*) An employee can only be assigned to work at her *available time*; available times are specified by the employee in advance.

(*ii*) On each day, an employee's time on duty has to be made up by consecutive time periods (together constituting a single *shift*).

(*iii*) An employee can be assigned to at most 5 shifts per week.

(*iv*) A shift has to be exactly 8 hours long for the full-timers, between 4 and 8 hours for the part-time and temporary-job employees and between 4 and 6 hours for the minors (underage employees).

(*v*) No employee can be assigned to more than 40 hours a week in total (note that this already comes from the conditions above). Moreover, full-time employees have to be given exactly 40 hours a week, and part-time employees at least 25.

Scheduling objectives. Many different objectives of staff scheduling are mentioned in the literature; in our case study, we restricted ourselves to two objectives that reflect the goals pursued in practice in our case study: (1) minimize the difference between the number of workers on duty and the staff demand in each time period, and (2) minimize the number of shifts. (The intuition behind the latter is that employees prefer few long shifts to many short ones.) In our model, these two objectives are aggregated using weights (or cost parameters) corresponding to their relative importance:

$\mathsf{c}^{\mathsf{under}}$ = the **c**ost per unit of **under**coverage (staff shortage) = 10.

$\mathsf{c}^{\mathsf{over}}$ = the **c**ost per unit of **over**coverage (staff excess) = 5.

$\mathsf{c}^{\mathsf{shift}}$ = the **c**ost per **shift** = 1.

Data inputs. There are two types of data supplied for a week's schedule: staff requirements for each time period, and employees' available times; the latter are specified in the form of a "from–to" range. The mathematical notation used in the rest of the paper is as follows:

re_{dt} = the number of staff **re**quired at time period t on day d.

$\underline{\mathsf{av}}_{ed}$ = the beginning of **av**ailable time for employee e on day d (the number of the first available period).

$\overline{\mathsf{av}}_{ed}$ = the end of **av**ailable time for employee e on day d (the number of the last available period).

These data are typically supplied in a spreadsheet, such as the one shown in Figure 1.

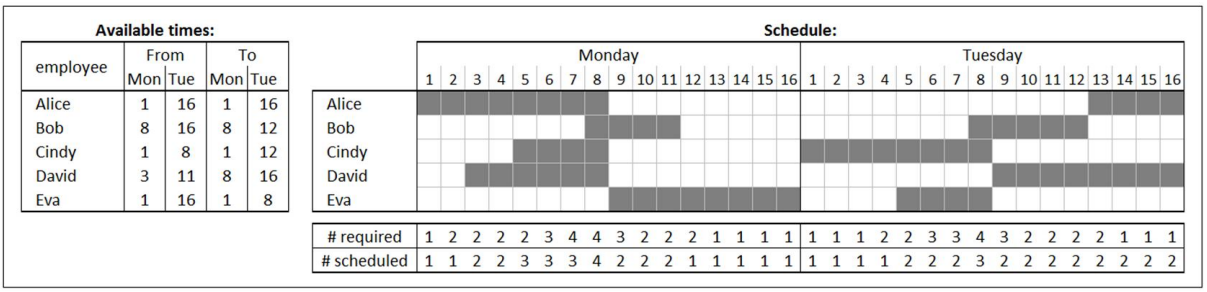


Figure 1: A minimalist example of a spreadsheet with data inputs and the resulting schedule. Data inputs are the numbers provided in the *Available times* table (the $\underline{\mathsf{av}}_{ed}$ and $\overline{\mathsf{av}}_{ed}$ values) and the *# required* row (re_{dt}). Using Model 1, the graphical schedule is created simply by applying suitable conditional formatting rules to the optimal 0–1 values of x_{edt} variables.

4 MILP formulations

In this section, we present two alternative MILP formulations of the problem, denoted as Model 1 and Model 2. The latter is analogous to the many variations of Dantzig's classical set-covering formulation [4]. Although this model is computationally quite efficient, it requires that large amounts of data are processed and prepared before the model is actually fed into an optimization software package, and the results are not easily translated into a readable schedule format, such as the one shown in Figure 1 (which is the format that was used by our study's restaurant in past). We explain these issues in detail in the discussion of Model 2; we nevertheless recognize that they might pose an obstacle for practical usage by small businesses. For this reason, we devised Model 1, which does not require any additional data processing—it works directly with the values of $\underline{\mathsf{av}}_{ed}, \overline{\mathsf{av}}_{ed}$, and re_{dt}—and, with proper conditional formatting in a *MS Excel* spreadsheet, the exported results look exactly the same as the schedule shown in Figure 1.

Model 1. Model 1 uses several decision variables, described in the list below. Binary variables are denoted by Latin letters, while Greek letters represent real non-negative variables.

δ^-_{dt} = staff shortage at time t, day d (negative part of the **d**eviation from the requirement re_{dt}).
δ^+_{dt} = staff overage at time t, day d (positive part of the **d**eviation from the requirement re_{dt}).
λ_{ed} = the **l**ength of the shift of employee e on day d ($= 0$ if e does not work on day d).
$x_{edt} = 1$ if employee e works in time period t on day d.
$y_{edt} = 1$ if employee e starts her shift in time period t on day d.
$z_{edt} = 1$ if employee e ends her shift in time period t on day d.

In order to make the MILP formulation easier to read, we use simplified notation for summation and iteration indices: unless stated otherwise, e, d, and t are indices that cycle through the entire sets E, D and T respectively, so that "$\sum_e$" effectively means "$\sum_{e \in E}$", "for all d" means "for all $d \in D$" etc. Moreover, we drop variable specification from the formulation, as we have already specified all variables above. The resulting formulation of Model 1 is as follows:

$$\text{minimize} \quad \mathsf{c}^{\mathsf{over}} \textstyle\sum_{d,t} \delta^+_{dt} + \mathsf{c}^{\mathsf{under}} \sum_{d,t} \delta^-_{dt} + \mathsf{c}^{\mathsf{shift}} \sum_{e,d,t} y_{edt}$$

$$\begin{aligned}
\text{subject to} \quad \mathsf{re}_{dt} - \textstyle\sum_e x_{edt} &= \delta^-_{dt} - \delta^+_{dt} && \text{for all } d, t, && (1a)\\
x_{edt} &= 0 && \text{for all } e, d, \text{ and } t < \underline{\mathsf{av}}_{ed} \text{ or } t > \overline{\mathsf{av}}_{ed}, && (1b)\\
x_{edt} - x_{ed,t-1} &\le y_{edt} && \text{for all } e, d, t, && (1c)\\
x_{edt} - x_{ed,t+1} &\le z_{edt} && \text{for all } e, d, t, && (1d)\\
\textstyle\sum_t y_{edt} &\le 1 && \text{for all } e, d, && (1e)\\
\textstyle\sum_t z_{edt} &\le 1 && \text{for all } e, d, && (1f)\\
\textstyle\sum_{d,t} y_{edt} &\le 5 && \text{for all } e, && (1g)\\
\textstyle\sum_t (t+1) z_{edt} - \sum_t t y_{edt} &= \lambda_{ed} && \text{for all } e, d, && (1h)\\
8 \le \lambda_{ed} &\le 16 && \text{for all } e, d, && (1i)\\
\lambda_{ed} &\le 6 && \text{for all } d \text{ and } e \in E_{\min}, && (1j)\\
\textstyle\sum_{d,t} x_{edt} &= 40 && \text{for all } d \text{ and } e \in E_{\text{full}}, && (1k)\\
\textstyle\sum_{d,t} x_{edt} &\ge 25 && \text{for all } d \text{ and } e \in E_{\text{part}}. && (1l)
\end{aligned}$$

The objective function is straightforward: the value of a schedule is a weighted sum of the total number of shifts and the deviations of scheduled staff numbers from the requirements; (1a) calculates these deviations from the schedule information contained in x_{edt}. Constraints (1b) through (1l) enforce the feasibility conditions (*i*) through (*v*) from section 2 in the following manner. (1b) requires that each person is assigned work only at their available times—which is (*i*). Constraints (1c) and (1d) provide the connection between x and y, z variables, saying that when x switches from 0 to 1 or vice versa in consecutive time periods, there has to be a beginning or an end of a shift, respectively; for the constraints to be correctly specified in the first and last time periods, we need to define constants $x_{ed0} = x_{ed,|T|+1} = 0$ for all d, e. (1e) and (1f) further require that each person starts and finishes a shift at most once per day—a one-shift-per-day requirement, which completes (*ii*). (1g) is a direct translation of (*iii*). (1h) calculates

the length of a shift (λ) from its beginning and end (y, z), and (1i), (1j) express the requirement regarding shift lengths in (iv). Finally, (1k) and (1l) together give (v).

Model 2. In Model 2, the main assignment variables do not assign employees to individual *time periods* as in Model 1, but to complete *shifts* instead. Obviously, the number of time periods per day and condition (iv) uniquely define the set of all possible shifts for each day, S. From the available times $\underline{\mathsf{av}}_{ed}, \overline{\mathsf{av}}_{ed}$, one can establish whether a particular shift can be assigned to a given employee on a given day—observing conditions (i) and (iv). Model 2 works with this sort of information, which has to be processed into the following parameters prior to solving the model:

$\mathsf{co}_{st} = 1$ if shift s **co**vers time period t.
$\mathsf{av}_{eds} = 1$ if employee e is **av**ailable for shift s on day d.
$\mathsf{le}_s =$ **le**ngth of shift s.

Note that the task of extracting these parameters from a data source such as the spreadsheet in Figure 1 is very tedious even for small-scale problems, if it is to be carried out manually. For instance, in our case study with 40 employees, 16 time periods and 7 working days, the array with co_{st} parameters has 6,336 entries and the array with av_{eds} has 52,920. Therefore, data preparation has to be automated somehow (for our case study, we programmed several VBA procedures in *MS Excel*).

Model 2 uses the same variables $\delta^-_{dt}, \delta^+_{dt}$ as Model 1; besides these, there is only one more type of binary variables: $x_{eds} = 1$ if employee e is assigned to shift s on day d. The formulation of Model 2 is given below; we use a similar convention regarding summation and iteration indices as with Model 1. The objective function is analogous to that of Model 1, as is the first constraint. Constraint (2b) enforces conditions (i) and (iv); (2c) and (2d) correspond directly to conditions (ii) and (iii) respectively, and constraints (2e), (2f) and (2g) together give (v):

$$\text{minimize} \quad \mathsf{c}^{\mathsf{under}} \textstyle\sum_{d,t} \delta^-_{dt} + \mathsf{c}^{\mathsf{over}} \sum_{d,t} \delta^+_{dt} + \mathsf{c}^{\mathsf{shift}} \sum_{p,s,d} x_{eds}$$

$$\begin{aligned}
\text{subject to} \quad \mathsf{re}_{dt} - \textstyle\sum_{s,e} \mathsf{co}_{st} x_{eds} &= \delta^-_{dt} - \delta^+_{dt} && \text{for all } d,t, && \text{(2a)}\\
x_{eds} &\le \mathsf{av}_{eds} && \text{for all } e,s,d, && \text{(2b)}\\
\textstyle\sum_{s} x_{eds} &\le 1 && \text{for all } e,d, && \text{(2c)}\\
\textstyle\sum_{s,d} x_{eds} &\le 5 && \text{for all } e, && \text{(2d)}\\
\textstyle\sum_{s,d} \mathsf{le}_s x_{eds} &\le 80 && \text{for all } e, && \text{(2e)}\\
\textstyle\sum_{s,d} \mathsf{le}_s x_{eds} &= 80 && \text{for all } e \in E_{\text{full}}, && \text{(2f)}\\
\textstyle\sum_{s,d} \mathsf{le}_s x_{eds} &\ge 50 && \text{for all } e \in E_{\text{part}}. && \text{(2g)}
\end{aligned}$$

Computational efficiency. Both models are of the MILP type with thousands of binary variables even for small-scale problems—which poses a potential threat to computational tractability; in Model 1, the number of binary variables is $3|E|\cdot|D|\cdot|T|$, and in Model 2 it is $|E|\cdot|D|\cdot|S|$. The latter number is typically greater, though the precise comparison depends on the number of time periods and the rules regarding acceptable shift lengths. In our case study, there were 26,880 and 52,920 binary variables in Model 1 and 2, respectively. On the other hand, Model 2 typically has fewer constraints, and more importantly, these constraints have a much simpler structure than those of Model 1. This simple structure seems to be effectively handled both in the pre-proccessing routines of today's MILP solvers and in the actual solving procedures. As a result, Model 2 outperformed Model 1 for all of our problem instances. To solve both models, we used the *Lingo 12* optimization software with its built-in solvers, installed on a PC with Intel dual dore 2.2 GHz processor and 2 GB RAM. Primarily, we worked with the empiric data from our case study (40 employees, 7 working days, 32 time periods); apart from that, we used several other test datasets with between 10 and 80 employees to analyze the effect of a varying problem size. In most cases, Model 2 was solved within 5 minutes; in a few isolated cases, the solver found a near-optimal solution within the first 5 minutes, and the optimal solution followed in another 5 to 10 minutes. With Model 1, the situation was hardly as optimistic. In terms of minutes, optimal solutions were found for none but the smallest problems (10 to 15 employees). For cca 15–25 employees, the optimal shedule was found in several hours' time, making it suitable for applications where the calculation can be carried out overnight. For larger problems, the solver was mostly unable to reach the optimum within first 10 hours; however, it typically did find a feasible schedule that was within 1–2% of the objective bound; in practical applications, these solutions can be used without too much loss of schedule quality.

5 Empiric vs. optimal schedules

In order to evaluate the scheduling efficiency in our restaurant, we compared the empiric and MILP-obtained solutions for a ten-week period in the winter and spring of 2012. On average, the empiric objective function was 3–5% above the optimal one. There are three important notes on this comparison. Firstly, for the reasons stated in §3, we only focused on day shifts and non-managers. This simplification probably improves the relative performance of empiric schedules; we believe that if we included long-term scheduling of night shifts and managers, the gap between empiric and optimal schedules would grew wider. Secondly, due to the presence of a large number of temporary jobs with loose schedule restrictions, the problem was not overly constrained, especially on working days, which again allowed the shedule managers to find quite good solutions by hand. We expect that in other settings, the empiric schedules would fall futher below the MILP-obtained ones. Thirdly, it should be noted that the objective function in MILP models does not capture all the pros and cons of various alternative schedules. For instance, it does not account for the objectives listed under (*d*) in §3; in empiric schedules, these criteria are typically reflected, at least to a certain extent. Improving on this is a task for our future work.

6 Conclusions

The major aim of this paper was to show that OR methods are under-exploited by Czech businesses—even such businesses as the fast food chains, which are mentioned as those setting an example of operations efficiency. In order to show this, we focused on a particular fast food restaurant's crew scheduling; here, staff is scheduled manually, which both incurs unnecessary scheduling cost (in our case study, scheduling effort amounts to nearly one man-day a week) and results in a sub-optimal schedule. Optimal schedules for this restaurant were found using MILP. We presented two MILP formulations: while Model 2 was adapted from the "classical" scheduling MILP models, Model 1 was our invention. Although the latter is computationally less efficient, it requires no additional data processing, and thus is easily implemented; we believe that it might find its use in small businesses. Finally, as mentioned in §5, future research is needed to incorporate additional criteria in the MILP models (such as grouping of work days, fairness etc.) and test the computational tractability of these extended models.

Acknowledgements

This research was supported by the Grant Agency of the Czech Republic, project no. GA402/12/P241.

References

[1] Brusco, M. J., Jacobs, L. W.: A simulated annealing approach to the solution of flexible labour scheduling problems. *The Journal of the Operational Research Society*, **44**(12) (1993), 1191–1200.

[2] Burns R. N., Koop, G. J.: A Modular Approach to Optimal Multiple-Shift Manpower Scheduling, *Operations Research*, **35**(1) (1987), 100–110.

[3] Card, D., Krueger, A. B.: Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply, *The American Economic Review*, **90**(5) (2000), 1397–1420.

[4] Dantzig, G.: A comment on Edie's traffic in toll booths, *Operations Research*, **2** (1954), 339–341.

[5] Ernst, A. T., Jiang, M., Krishnamoorthy, M., Sier, D.: Staff scheduling and rostering: A review, *European Journal of Operational Research*, **153** (2004), 3–27.

[6] Glover, F., McMillan, C.: The general employee scheduling problem: An integration of MS and AI. *Computers & Operations Research*, **13**(5) (1986), 563–573.

[7] Love R. R., Hoey J. M.: Management Science Improves Fast-Food Operations, *Interfaces*, **20**(2) (1990), 21–29.

[8] Tanomaru, J.: Staff scheduling by a genetic algorithm with heuristic operators. In: *Proceedings of the 1995 IEEE International Conference on Evolutionary Computation*, (1995), 456—461.

Impact of the real exchange rate on Czech trade

Libor Žídek[1], Daniel Němec[2]

Abstract. Since the beginning of the transformation process in 1990, the Czech crown has operated in several different exchange rate regimes. The Czech currency appreciated in real terms in all of the regimes. The cause was higher inflation in the Czech economy compared to inflation rates in the trading partners' economies during the periods of the fixed regimes. Later on, the main cause was in nominal appreciation of the crown. It is surprising that in the meantime Czech exports were increasing enormously (in absolute terms as well as percentage of GDP). The goal of the article will be to analyse and explain the seemingly positive impact of real exchange rate on the development of Czech foreign trade. The cointegrating (long-run) relations are estimated using Johansen's multivariate procedure. Estimates of the short-run dynamics are obtained using the error-correction techniques.

Keywords: real exchange rate, foreign trade, Czech economy

JEL Classification: F31, C22
AMS Classification: 91B84

1 Introduction

The development of the Czech economy since the end of the centrally planned period is full of interesting aspects. We would like to concentrate on changes in the real exchange rate and Czech international trade (exports). These two highly important measures of any economy developed against economic intuition in the Czech case. The real exchange was appreciating continuously for nearly the whole period. We would expect that this appreciation should be negative to the growth of Czech exports. The economy opened and from a highly closed economy became a highly open one. Both sides of the trade – exports as well as imports increased in absolute as well as in relative terms. Strong growth of exports is development that we would not expect from the strong appreciation of the real exchange rate. The goal of the article is to explain the seemingly positive impact of real exchange rate on the development of Czech foreign trade.

2 Development of the real exchange rate and trade

Development of the real exchange rate was highly interesting during the whole period. For different reasons the real exchange rate appreciated since 1991. We can see two main periods of development of the nominal exchange rate. The first period of fixed exchange rate lasted till 1997 and it was followed by the second period of (nearly) clean floating. Reasons for real appreciation of the exchange rate were different but the currency appreciated in real terms for the whole period.

Period of fixed exchange rate

The exchange rate system was highly distorted during the communist reign as well as other aspects of the functioning of the economy. Foremost, there was a system of multiple exchange rates. Czechoslovak production was not able to assert itself on the international (western) markets because of the poor quality of production. The multiple exchange rates were used to maximize chances to get hard (convertible) currencies into the country. There was an additional system of surcharges with the same intention (Jirges and Plchová [6]). Winiecki [11] shows that the official exchange rate was far from the unofficial exchange rate on the black market. This situation changed already in the first year of the transformation – in 1990.The government realized that the nominal exchange rate was crucial for the whole transformation process. Several important decisions were made. It was decided to unify the system of multiple exchange rates first of all. At the same time, there were three large devaluations during the year and the crown lost 100 % in total against the dollar. The government decided that the crown should be functioning in the fix exchange rate regime. And it was decided that the exchange rate was to be fixed to a basket of five currencies. The exchange rate was supposed to be an anchor of the whole system – one stable value in the volatile transformation environment. It meant at the same time that the setting of the exchange

[1] Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, l_zidek@centrum.cz.
[2] Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, nemecd@econ.muni.cz.

rate was a highly difficult task – Klaus says that it was the most difficult decision in the whole transformation process. The crown was, at the end of 1990 (after the devaluations), undervalued. It created something that was termed as the first transformation cushion or pillow. Generally, it was a temporary help for Czechoslovak exporters. But it was only temporary precisely because of the following real appreciation of the crown. From the beginning of the year 1991 the crown became internally convertible – as a consequence transactions on the current account were allowed without any restrictions. It had straight consequences for the development of Czechoslovak trade, which we will mention later.

The above mentioned points were crucial changes of the exchange rate system that took place at the beginning of the transformation process. Later on, it was decided to apply a small fluctuation zone ± 0.5 % and the basket was reduced to just two currencies – the dollar and the D-mark in 1993. In 1996, the fluctuation zone was widened to ± 7.5 %. This regime ended abruptly with currency crisis in May of 1997 – we will describe it in the following subsection. What happened meanwhile with the real exchange rate? The very first step in the form of devaluation of the nominal exchange rate led obviously to depreciation of the real exchange rate as well. But since the beginning of 1991 the nominal exchange rate was de facto fixed. At the same time, the price level in Czechoslovakia and later on in the Czech Republic highly increased in comparison to the main trading partners because of price liberalization, following price deregulations, relatively high inflation expectations and the tax reform. All of these steps led to a relatively high inflation that meant continuous strengthening of the real exchange rate. Frait and Kotlán [4] show that both Hungary and Poland devaluated their currencies and used crawling pegs in some periods during the 1990s. As a consequence, their real exchange rates did not appreciate as much as in the Czech Republic.

Period of floating exchange rate

The Czechoslovak economy suffered from currency crisis in the spring of 1997. The central bank tried to defend the fix exchange rate and sell 3 billion dollars from its reserves on it but it had to abandon its efforts in the end. Consequently, it changed its main policy from the fixed exchange rate regime to inflation targeting. Meanwhile, the exchange rate was altered to managed floating. In reality, the central bank intervened only rarely in the following years. It means that the system was close to clean floating. The Czech exchange rate followed the path of continuous nominal appreciation in the following years. At the same time, the central bank was able to decrease the inflation level in the country after applying inflation targeting. Inflation declined to levels comparable to levels in the European Union or even lower after the turn of the century. The consequence was that the process of real appreciation continued but the main cause was different. The engine of the appreciation was the nominal appreciation of the crown instead of increasing of the price level as it had happened in the first transformation decade. To sum it up, the real exchange rate was first of all in 1990 depreciated with deep nominal depreciation of the crown. But then the process of real appreciation was permanent. There was only difference in the main cause of this process. The cause changed from growth of price level to appreciation in the nominal exchange rate after the currency crisis in 1997. The overall development of the real exchange rate after 1994 can be seen in the Figure 1. We should notice the continuous growing trend for the entire period.

3 Development of the international trade (exports)

What happened meanwhile with Czech trade? We could expect that strong and continuous real appreciation should have negative influence on exports. But the reality was entirely different. In this section we analyze the development of Czech trade after 1990 with emphasis on exports. We generally concentrate on trade with Western countries. At the end of the communist regime the economy was relatively closed – exports to GDP can be estimated to 20 %. The number is low if we consider the small size of the Czechoslovak economy at that time. Most of the trade was focused on the Eastern markets – 70 % (Půlpán, [10]). And exports to the West generally consisted of goods with low added value. The trade was distorted and planned as other aspects of the Czechoslovak economic life at that time. The government after 1990 believed that competition was the only way to improve efficiency of the large monopolies and thus the entire Czechoslovak economy. It was difficult to start a new business from scratch and compete with these large companies. The government for this reason decided to open the inner market for foreign competition already at the beginning of 1991. It applied inner convertibility of the crown and decreased tariffs. Inner convertibility allowed exports of anything by anyone (there was an obligation for importers to sell the hard currency that they obtain to the central bank at the same time). On the other hand, the government was deeply worried about the exchange rate. Understandably, high foreign currency reserves were needed to defend the fixed exchange rate. And the government did not know how the markets would react to the general opening of the economy. For this reason, temporary 20 % import surcharge was applied since January of 1991 but it was step by step decreased to zero in the following 2 years. The Czechoslovak economy had generally highly liberalized trade (both exports as well as imports) since the very beginning of the transformation process. We should mention that there were negative circumstances in the outer relationship at the begin-

ning of 1990s. Foremost the Eastern markets quickly and totally disintegrated. The economy lost its main trading partners from the previous decades and Czech exports with these territories slumped to one fifth of the previous level (Jonáš, [7]). On the other hand, the trade quickly re-orientated towards the Western markets. The economy became integrated into the world economy and its trade started to grow.

There were possible boosts for exports in the following years in the forms of membership in international organizations – foremost the WTO and later on the EU and formally in splitting of the Czechoslovakia. This probably helped to endure the growth process of Czech. But generally, the impact of these changes was relatively small. The growth was enormous. The level of exports to GDP increased from 25 % at the end of the 1980s to close to 70 % before the economic crisis twenty years later. The Czech economy deeply integrated into the world economy in a relatively short time period. At the same time, other characteristics of the trade show that territorial orientation switched from the Easter to the demanding Western markets (foremost the EU). And commodity structure of exports returned to machinery and vehicles – (goods with a relatively higher added value) after a short run slump in exports of more sophisticated goods at the beginning of the first transformation decade. From these points of views we can see the development of Czech exports as one of the greatest achievements of the whole transformation process. The consequence is that the quality and efficiency of the Czech production had to strikingly increase during the period. And in the context of our article we should emphasize that it was achieved under the conditions of ever growing real exchange rate.

4 Econometric models and estimation results

The main goal of this contribution is to analyse and explain the seemingly positive impact of real exchange rate on the development of Czech foreign trade. The lack of plausible and consistent data sets allows us to analyse only two exchange rate regimes: a pegging system with a band (in 1996 and 1997) and the regime of managed floating (since 1997). Our econometric approach is straightforward. We carry out tests of cointegration (Johansen's approach) and we estimate long-run relationships among cointegrated variables and short-term dynamics using error correction approach. Finally, we determine relative importance of all relevant explanatory variables on the foreign trade. But, before doing that we need to obtain operational measures of all these variables. The data set used for estimation is from the first quarter 1996 to the fourth quarter 2011. The data comes from the OECD database, from the database ARAD of the Czech National Bank and from the Eurostat. The observed variables (for the Czech Republic and EU27) are as follows: real effective exchange rate (Czech crown): deflated by CPI, foreign trade weights of 2010, index 2010=100 (ARAD); Export and import prices: index 2000=100 (OECD); Real consumption, government spending, investment, exports, and imports: millions of national currency, chained volume estimates, national reference year, quarterly levels, s.a. (OECD). Using these data sets, we have constructed appropriate model variables which are presented in Table 2 (including their stationarity properties). All variables are in logarithms. In addition, we created three dummy variables for periods of pegging system with a band (1996-1997), for periods of EU membership of Czech Republic (3rd quarter 2004 and later) and for periods of the last economic slowdown (from the 4th quarter 2008 to 4th quarter 2011). But, these variables do not influence our results (the corresponding model variables prove to be statistically insignificant in both the cointegration relationships and the corresponding error-correction representations).

In our contribution we use a time varying measure of real exchange rate volatility. This proxy variable is constructed by the moving-sample standard deviation which is similar to those employed in the international trade literature (see Arize et al. [1]):

$$J_{t,m} = \left[\frac{1}{m} \sum_{i=1}^{m} \left(R_{t+i-1} - R_{t+i-2} \right)^2 \right]^{1/2}, \tag{1}$$

where R is the natural logarithm of real effective exchange rate and $m = 7$ is the order of the moving average. In our case, we use a centered moving average, i.e. $i = -(m-1)/2, \ldots, (m-1)/2$. We have computed real exchange rate volatility using a standard GARCH in mean model applied on differenced real exchange rate variables (using Bayesian information criteria, a GARCH(1,1) model was selected) as well (this approach may be found in Omojmite and Akpokodje [9]). It may be shown that the resulting trajectories have similar dynamic (only the levels of these variables differ).

There are many approaches (models) to estimate equilibrium real exchange rate, e.g. behavioural equilibrium exchange rate mode (BEER) or permanent equilibrium exchange rate model (PEER). Both model frameworks involve estimating of the equilibrium exchange rate using cointegration analysis. Bęza-Bojanowska and MacDonald [2] estimates real Polish zloty/euro equilibrium rate using these approaches. Frait and Komárek [5] presented and estimated real equilibrium exchange rate for the Czech crown. Both papers are focused on the driving

forces standing behind the evolution of the real exchange rate and its equilibrium trajectory. Our contribution is intended to be focused on analysis of the determinants of the Czech foreign trade (in particular of the exports). To estimate real equilibrium exchange rate, we use Hodrick-Prescott filter with the standard smoothing parameter (1600) for the quarterly data. This filter is applied on the real exchange rate time series. This approach is sufficient to approximate the development of the equilibrium real exchange rate. We are aware of the fact that we are not able to fully understand the set of factors that possibly caused the real appreciation in the past. But, these factors are explained by Frait and Komárek [5]. It should be noted that this variable was used especially for robustness checking (i.e. alternative models specifications).

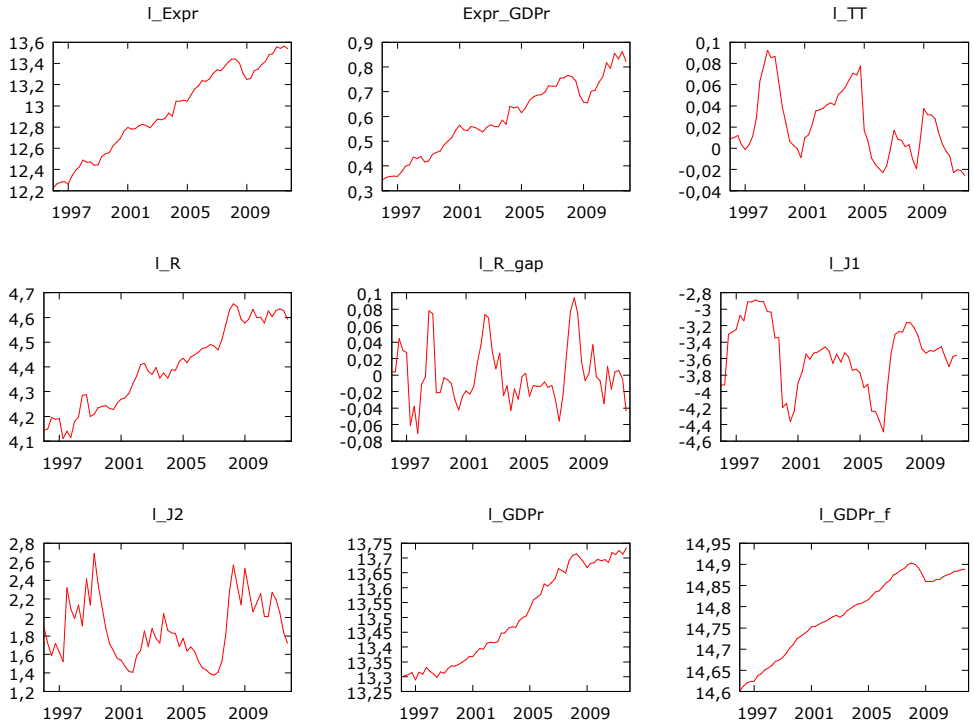


Figure 1 Model data

Determining the importance of explanatory variables is of practical relevance. This topic is a controversial one and there is no a unique measure. Commonly used measures are: zero-order correlations, partial correlations, semiparametrical correlation, standardized regression coefficients, Akaike weights and independent effects (see Murray and Conner [8] for further details and Monte Carlo simulations). In our evaluation of relative importance we use the concept of standardized coefficients and reduction in R^2 statistics (see Bing [3]). Standardized regression coefficients are identical to the estimates obtained from the regression on the standardized variables.

Our analysis of possible cointegration starts with the determination of the stationarity properties of our time series. If the series do not follow the same order of integration, then there can be no meaningful long-term relationships among them and we should concern only on the short-term dynamics. The Augmented Dickey-Fuller (ADF) test has been used to detect unit roots in the data. The results of unit-root tests are presented in Table 1. We can see that only the logarithms of real exports, real GDP of the Czech Republic and real GDP of the EU27 may be treated as integrated of order 1. The lag length is based on the significance of the most lagged variable where the residuals are white noise. One asterisk implies that we can reject the null hypothesis that the time series has a stochastic trend or contains a unit root at 10% significance level, two asterisks implies 5% level of significance and three asterisks 1% significance level. ADF tests have been carried on including constant (c) and constant and trend (c, t) in the corresponding test equations. We are able to reject the null hypothesis that the differenced time series have a stochastic trend or contain unit root at 1 and 5% significance level respectively. Only the variables of real exports, Czech real GDP and real GDP of the EU27 contain the unit root. Real exchange rate seems to be a trend stationary variable. All the model data are depicted in Figure 1. As mentioned previously, the cointegrating (long-run) relations are estimated using Johansen's multivariate procedure. Estimates of the short-run dynamics are then obtained using the error-correction technique. We have constructed many models (including interaction terms with our dummies and various volatility measures) but the number of cointegration vectors and the effects estimated coefficients remained robust. Table 3 presents the results for two models: Model 1 includes four endogenous variables (real exports, real GDP, real exchange rate, foreign GDP),

unrestricted constant, restricted trend and two unrestricted stationary variables in error-correction equation (terms of trade and exchange rate volatility); Model 2 treats the foreign GDP as an exogenous (restricted) variable in the cointegration equation. In both models, only one cointegration relationship has been identified (using 5% level of significance). Optimal lag length (order 2) was selected using the information criteria.

Variable	Notation	ADF (c) Level	ADF (c, t) level	ADF (c) difference	ADF (c, t) Difference
Real exports	l_Expr	-1.14	-2.00	-6.97***	-6.97***
Real exports to GDP ratio	Expr_GDPr	-0.87	-3.92**	-3.93***	-3.90**
Terms of trade	l_TT	-3.18**	-3.61**	-4.46***	-4.49***
Real exchange rate	l_R	-0.81	-4.21***	-6.29***	-6.24***
Real exchange rate gap	l_R_gap	-5.09***	-5.03***	-6.63***	-6.57***
Real exchange rate volatility (moving average)	l_J1	-3.27**	-3.21*	-4.07***	-4.03***
Real exchange rate volatility (GARCH)	l_J2	-2.81*	-2.8	-9.59***	-9.52***
Real GDP	l_GDPr	-0.48	-1.68	-3.34**	-3.25*
Real GDP for EU27	l_GDPr_f	-2.19	-1.04	-4.36***	-4.88***

Table 1 List of model variables (in logarithms) and tests of stationarity

Rank	Trace statistics Model 1	Max. eigen. statistics Model 1	Trace statistics Model 2	Max. eigen. statistics Model 2
0	77.52 (0.002)	36,63 (0.010)	71.80 (0.000)	42.96 (0.000)
1	40.90 (0.077)	20.98 (0.198)	28.84 (0.085)	19.11 (0.052)
2	19.91 (0.235)	14.33 (0.241)	9.727 (0.303)	9.73 (0.143)
3	5.58 (0.524)	5.58 (0.526)	-	-

Table 2 Johansen's cointegration test (p-values in parenthesis)

	l_Expr	l_R	l_GDPr	l_GDPr_f	Trend
Model 1	1.000	1.620 (0.254)	-0.324 (0.313)	-1.768 (0.370)	-0.024 (0.003)
Model 2	1.000	0.957 (0.187)	-0.123 (0.234)	-2.351 (0.276)	-0.018 (0.003)

Table 3 Long-term cointegration vectors (normalized, standard deviations in parenthesis)

	Model 1				Model 2		
Variable	d_l_Expr	d_l_R	d_l_GDPr	d_l_GDPr_f	d_l_Expr	d_l_R	d_l_GDPr
const	-4.757***	-2.120**	-0.644	-0.570***	-12.6***	-2.700	-2.347***
d_l_Expr (-1)	-0.178	-0.065	0.039	-0.007	0.047	0.044	0.051
d_l_R (-1)	-0.157	0.247	0.180**	0.080***	0.157	0.104	0.211***
d_l_GDPr (-1)	0.228	0.660*	-0.212	0.044	0.230	0.603*	-0.212
d_l_GDPr_f (-1)	1.984**	0.635	0.088	0.380***	-	-	-
l_J1	0.009	0.0133	-0.006	0.000	0.025**	0.012	-002
l_TT	-0.305**	-0.074	-0.033	-0.010	-0.479***	-0.036	-0.085
EC1	-0.414***	-0.187**	-0.054	-0.049***	-0.632***	-0.136	-0.117***
R^2	0.55	0.17	0.22	0.47	0.60	0.10	0.29

Table 4 VECM models (*, **, *** statistical significance of parameters at 1, 5 and 10% significance level)

We have found only one cointegration vector using Johansen cointegration test (see Table 2). Conditioned by these results, the final VECMs are estimated. Table 4 contains the estimates of cointegration equations. Table 4 contains the estimates of short-term dynamics. Long-term relationship among variables shows that the exports are negatively related to the development of the real exchange rate. But, this negative connection is neglected by the trend behaviour of the exports (average quarterly growth is 2.4%). An important factor for determination of the Czech foreign trade is the foreign demand (long-term elasticity is 1.77%). As for the short-term dynamics, changes in foreign demand a terms of trade play an important role. Error correction term (EC1) is significant and

the adjustment coefficients suggest that the export dynamics is strongly determined by the development of all variables in the cointegration equation. The influence of volatility is ambiguous. Table 5 shows the estimates of VECMs using standardized variables. Unfortunately, it is not possible to standardize the variables in the cointegration regression because all the variables are nonstationary. We are focusing on the export equation only. The row denoted "sign. only" means, that standardized coefficient were computed only using a regression omitting insignificant variables from the original VECM model, the row "R^2 change" shows the change of R^2 in the regression omitting the particular variable. It is evident that the most important factors influencing the development of the exports are terms of trade and error correction term. This term contains the real exchange rate variable and the foreign demand variable. We can conclude (regarding the elasticity estimates from the cointegration vector) that the influence of both these variables is similar. But, this conclusion should be interpreted in a way that the permanent changes of real exchange rate and foreign demand matter. Especially in the case of the estimated expected quarterly trend growth of 2.4% influenced by the factors discussed by Frait and Komárek [5].

	d_l_Expr(-1)	**d_l_R (-1)**	**d_l_GDPr (-1)**	**d_l_GDPr_f (-1)**	**l_J1**	**l_TT**	**EC1**
Model 1	-0.177	0.132	0.086	0.292	0.090	-0.251	-0.753
R^2 change	0.02	0.008	0.005	0.059	0.019	0.039	0.250
sign. only	-	-	-	0.259	-	-0.144	-0.632
Model 2	0.106	0.189	0.112	-	0.251	-0.390	-0.916
R^2 change	0.008	0.017	0.009	-	0.038	0.082	0.339
sign. only	-	-	-	-	0.131	-0.326	-0.724

Table 5 Relative importance - standardized coefficients, change of R^2

5 Conclusion

In our contribution, we have found the long-term relationship among real exports, real exchange rate and the foreign demand. We have shown that the real exchange rate and its dynamics have a negative impact on the real exports and its development. Seemingly positive impact of real exchange rate on the development of Czech foreign trade was not proved. Czech GDP does not affect the exports both in the long-run and in the short-run. In the short-run, the export dynamics is influenced by the terms of trade and by the volatility of real exchange rate (this influence is ambiguous). The most important factor in determining the dynamics of the export are the deviations from the export equilibrium which may be treated as a desired volume of exports.

Acknowledgements

This work is supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/0786/2011.

References

[1] Arize, A. C., Osang, T., and Slottje, D. J.: Exchange-Rate Volatility and Foreign Trade Evidence from Thirteen LDC's. *Journal of Business & Economic Statistics* **18** (2000), 10–17.
[2] Bęza-Bojanowska, J., and MacDonald, R.: The Behavioural Zloty/Euro Equilibrium Exchange Rate. National Bank of Poland Working Paper No. 55, 2009.
[3] Bring, J.: How to Standardize Regression Coefficients. *The American Statistician* **48** (1994), 209–213.
[4] Frait, J., and Kotlán, V.: Macroeconomic stabilization and macroeconomic policy. In: *Politics and economics in the Czech transition* (Cueto, C.de, Hájek, L. eds.). Gaudeamus, 2000.
[5] Frait, J., and Komárek, L.: Real exchange rate trends in transitional countries. Warwick Economic Research Paper No. 596, 2001.
[6] Jirges, T., and Plchová, B.: *Zahraniční obchod a národní ekonomika*, VŠE, Praha, 1996.
[7] Jonáš, J.: *Ekonomická transformace v České republice*. Management Press, Praha, 1997.
[8] Murray, K., and Conner, M. M.: Methods to quantify variable importance: implications for the analysis of noisy ecological data. *Ecology* **90** (2009), 348–355.
[9] Omojimite, B. U., and Akpokodje, G.: A Comparative Analysis of the Effect of Exchange Rate Volatility on Exports in the CFA and Non-CFA Countries of Africe. *Journal of Social Sciences* **24** (2010), 23–31.
[10] Půlpán, K.: *Nástin českých a československých hospodářských dějin do roku 1990*. Institut ekonomických věd, Praha, 1993.
[11] Winiecki, J.: Regional survey. In: *Macroeconomics of Transition* (Winiecki, J., Kondratowicz, A. eds). Routledge, London and New York, 1993.

Cluster analysis of the Liberec region municipalities

Miroslav Žižka[1]

Abstract. The article deals with the cluster analysis of municipalities in the Liberec Region. It builds on the results of factor analysis, which defined seven significant factors that characterize the socio-economic status of communities. The input for cluster analysis was the factor loadings, which allowed making multi-dimensional classification of the studied municipalities. To build the structure of clusters there were used both hierarchical clustering methods and non-hierarchical clustering methods with the help of k-means. In the first case the Ward's method was utilized to create clusters and the results were graphically displayed by tree diagram. With the help of the division procedure an initial cluster was divided into 2 to 4 smaller clusters. Two basic clusters can be classified as urban and rural. Urban cluster is characterized by better features of population settlement, age structure, civic amenities and the branch structure. On the contrary, cluster consisting of rural communities has an aging population, population migrating to cities, poor civic amenities and the employment structure is dominated by manufacturing and agriculture. When lowering the clustering level, it can be found that rural cluster breaks down to agricultural, submountaneous and cross-border areas. In the second stage, the method of non-hierarchic clustering using k-means was applied on the results of factor analysis. A set of municipalities of the Liberec Region was gradually divided into 2 to 10 clusters, which were analyzed in detail using the R-square index and Calinski-Habarasz F index. Taking into account the requirement that the clusters were not formed only by a few elements and outlying values, the region was finally divided into 6 clusters, which can be classified as urban, micro-regional, economically weak and rural focused on housing, services and agriculture.

Keywords: Cluster analysis, hierarchical cluster analysis, non-hierarchical cluster analysis, factor loadings, Ward's method, k-means clustering, measure of disagreement, recall coefficient, F-ratio, Calinski-Habarasz F index.

JEL Classification: C38, R11, R15
AMS Classification: 62H30, 91C20

1 Introduction

Cluster analysis is one of the frequently used multivariate statistical methods in social sciences. The examples of practical application of cluster analysis include classification of the EU regions in terms of business environment on the basis of factor loadings of the individual regions [6], the division of countries in Europe, the Middle East and Africa in similar groups in terms of the position of high tech companies, the economic environment of the country and its competitiveness index [4] or using a hierarchical clustering method to classify SMEs in terms of their ICT competencies [1]. Another example is the creation of a number of client profiles of bank customers with the help of the modified hierarchical method using a CF tree [2].

In the article [9], a methodology for evaluating the socio-economic status of municipalities in the Czech Republic was characterized. Based on this methodology, seven significant factors (F1 - unemployment, F2 - domicile attractiveness, F3 - population settlement, F4 - age structure, F5 - civic amenities, F6 - branch structure and F7 - economic activity; below are used only abbreviated designation of factors F1 to F7) have been identified, including the factor loading of the individual municipalities.

This article aims to build on the results of previous research and divide communities characterized by means of factor loadings into homogeneous clusters, if possible, in which the individual municipalities will be as similar as possible in terms of factor loadings. Individual clusters give an idea of socio-economic status of the cluster while allowing a classification of municipalities surveyed with regard to individual factors of the socio-economic environment. In the first stage, methods of hierarchical clustering were used to construct clusters, in the second stage; they were followed by methods of non-hierarchical clustering.

[1] Technical University of Liberec/Faculty of Economics, Department of Business Administration, Voroněžská 13, 461 17 Liberec, Czech Republic, miroslav.zizka@tul.cz.

2 Divisional hierarchical clustering

In this procedure, we assume that all municipalities form a cluster and by its gradual division we obtain a larger number of clusters so that we would end up in each municipality. In fact, we will attempt to divide a basic set of the Liberec Region municipalities into several large, internally homogeneous clusters that will vary in individual factors of the socio-economic environment.

Similarity of municipalities was evaluated first by tree diagram of objects; the distance of objects was measured by the squared Euclidean distance which forms the basis of Ward's clustering method (for example, see [5]). Ward's method was used to construct clusters because graphical analysis showed that the surveyed municipalities tend to be grouped into multi-dimensional ellipsoids. One problem was the fact that there were 215 municipalities and it did not allow a clear description of their names on the vertical axis of tree diagram (see Figure 1). For the assignment of individual municipalities in clusters was therefore used the table of amalgamation schedule in the STATISTICA CZ 10 programme. Clustering level h = 210 split the set of municipalities into 2 unequally large clusters. The first cluster includes 46 municipalities and is made up of larger regional settlement (the average number of inhabitants is 7,454); while the second cluster contains 169 mostly small municipalities (average number of inhabitants in this cluster is 569). We can say that on the significance level $\alpha = 5\%$ difference in the average number of inhabitants in the two clusters is statistically significant (p-value < 0.0001).

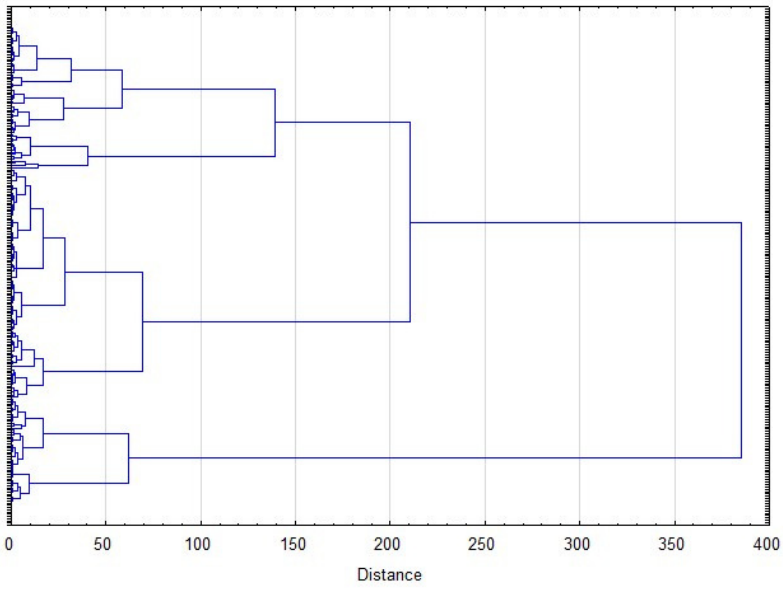


Figure 1 Tree Diagram

Ward's Method, Squared Euclidean distance

When comparison was made of the averages of factor loadings for both clusters (see Figure 2) at significance level $\alpha = 5\%$ substantial differences were verified between clusters of factors F3 (p-value < 0.0001), F4 (p-value = 0.0453), F5 (p-value < 0.0001) and F6 (p-value < 0.0001). On the contrary, a statistically significant difference between the average of factor loadings were not detected in clusters of factors F1 (p-value = 0.2989), F2 (p-value = 0.6800) and F7 (p-value = 0.9015). Based on these findings, we can say that cluster 1 is characterized by better parameters of indicators relating to population settlement, age structure, civic amenities and the branch structure. Due to the fact that this cluster usually includes cities and larger municipalities, the results are logical. As shown in the atlas [8], larger communities are characterized by higher population density, usually a positive migration balance, are equipped with basic educational and medical facilities, and most of the population is economically active in the tertiary sector. On the contrary, cluster 2 is composed of mostly small, rural communities, often facing an aging population and outflow of inhabitants to the cities, there is a need to commute for civic amenities and manufacturing industry and agriculture play more important role in the employment structure.

With the decrease of clustering level at h = 140, we obtain three clusters of municipalities which include 46, 105 and 64 municipalities. The first cluster of municipalities yet remains the same (see bottom of Figure 1). The second cluster now includes a municipality with an average population of 474 and the third cluster has an average size of 724 residents, but in both cases with a high degree of variability (v = 0.67, 1.15 respectively). Cluster 2 is represented by rather agricultural areas of the region, while cluster 3 includes the submountaneous and cross-border areas of the region. If we reduce the clustering level arbitrarily at h = 70, we obtain 4 clusters of municipalities with 46, 105, 16 and 48 municipalities each. In an analogous way, we can continue to form other,

smaller clusters of municipalities. In principle, this hierarchical approach provided us with the basic information about the structure of clusters of municipalities in the Liberec Region and for further analysis we will use the method of non-hierarchic clustering.

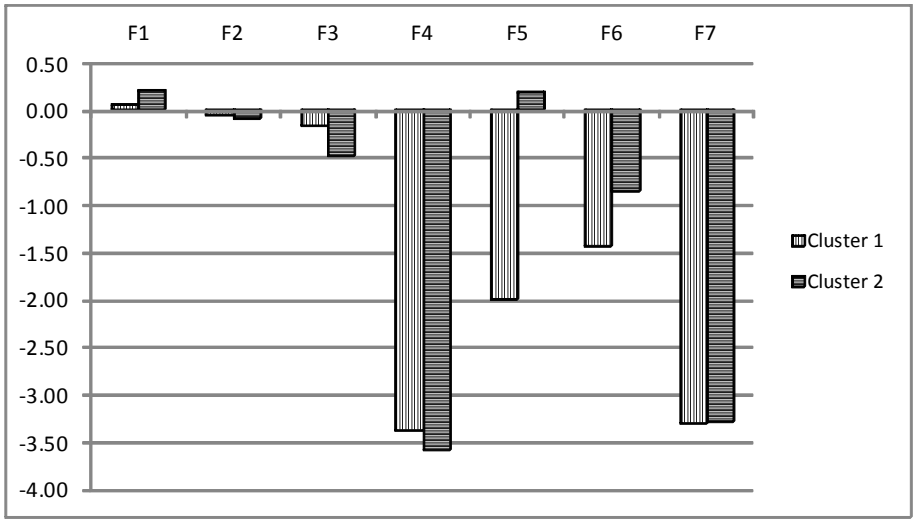


Figure 2 Average values of factor loadings in clusters 1 and 2

3 Clustering by the k-means method

The aim of the application of this method is the division of individual municipalities of the Liberec Region in the predetermined number of clusters. The algorithm is described, for example, in the work [3]. The initial decomposition can be determined randomly, in our case, however, it can be based both on the previous result of hierarchical clustering, where "reasonable" number of clusters appears to be 2 to 4, and on the administrative structure of the Liberec Region that is divided in 4 districts and 10 administrative districts of municipalities with extended powers. For these reasons, the division of municipalities in the region was tested in 2 to 10 clusters. Calculations were again performed in the STATISTICA CZ 10 programme. Three options to determine the initial cluster centers are available: maximizing the initial distance between clusters, ordering observations according to distance and choice of objects in constant intervals or selection of the first N observations. Individual variants were compared using analysis of variance to determine whether the averages of the individual factors are significantly different between groups. For the option k = 2 clusters, the least suitable method appeared to be arrangement according to distance (at significance level $\alpha = 5\%$ the null hypothesis of conformity of variances was not rejected in the case of 4 factors, in other two methods, H0 was not consistently rejected in one variable). For the option k = 3 clusters, H0 was not rejected consistently in the case of one factor in the following methods - sort distances and choose the first N observations. In the case of k = 4 and a larger number of clusters, it can be already said that congruence of variance was not verified in any factor (see Table 1), and for all methods. Table 2 shows the number of municipalities across clusters.

The determination of "optimal" number of clusters is a highly discussed question of cluster analysis. As stated in [5], there is no objective way to determine such a termination criterion. In the literature, (see for example [5], [7]), there is therefore described a number of different criteria and methods for assessing the quality of clusters.

Factor	between SS	Df	within SS	Df	F	p-value
F1	97.1268	3	66.97478	211	101.9974	0.000000
F2	3.1079	3	57.59572	211	3.7952	0.011119
F3	4.8871	3	46.76894	211	7.3494	0.000104
F4	20.1172	3	51.18085	211	27.6453	0.000000
F5	192.8578	3	33.22964	211	408.1998	0.000000
F6	23.8639	3	30.18421	211	55.6060	0.000000
F7	44.8476	3	27.92743	211	112.9457	0.000000

Table 1 Analysis of variance

Notes: N = 4 clusters, method of maximizing of initial distances between clusters

One of the simpler methods for evaluating clusters is the **measure of disagreement *MD*** (1), based on confusion matrix [7]. Suppose that we know in advance the structure of *P* division of municipalities in clusters and further the structure *C* obtained by cluster analysis. Confusion matrix contains the number of objects occurring simultaneously in cluster structure *P* and cluster structure *C*. Number of common objects shall be denoted as n_{hh}. The number of clusters in *C* and *P* is the same. The structure of clusters known in advance was derived from the administrative structure of the Liberec Region, i.e. 4 districts and 10 administrative districts of municipalities with extended power (MEP).

$$MD = \frac{n - \sum_{h=1}^{k} n_{hh}}{n} \tag{1}$$

When comparing the composition of the municipalities in 4 districts with the structure of 4 clusters obtained by cluster analysis, we find that MD = 0.56. For 10 MEP and 10 clusters, the degree of disagreement is even higher, namely 0.72. Therefore, it can be said that there are clusters of municipalities across districts and MEP which do not copy the existing administrative structure of the region that much.

Other measures based on a comparison with the predetermined classification are the **accuracy coefficient** $P_{hh'}$, expressing a share of common objects of two clusters on the number of objects from the structure *C* or the **recall coefficient** $R_{hh'}$, indicating a share of common objects on the number of objects of the cluster structure *P*. Combining the two above-mentioned measures we receive the **F-ratio** (2), which is their harmonic average. The result is a value from 0 to 1.

$$F_{hh'} = 2\frac{P_{hh'}R_{hh'}}{P_{hh'} + R_{hh'}} \tag{2}$$

In the case of F-ratio and 4 clusters, cluster 4 (0.61) shows the highest value where there is a high number of common municipalities in the Semily district, while cluster 2 consists of municipalities from all districts (F = 0.17). If we compare the structure of the emerged 10 clusters with the structure of 10 MEP, then the highest congruence was observed in cluster 9 (F = 0.32), where there are most of the municipalities from Turnov MEP.

Cluster No.	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10
1	54	70	68	1	1	15	2
2	13	46	15	15	39	4	22
3	39	35	39	36	4	40	16
4	109	12	35	79	18	26	3
5		52	12	33	71	1	33
6			46	39	39	38	1
7				12	15	52	15
8					28	2	38
9						37	26
10							59

Table 2 Number of municipalities in individual clusters

To assess the number of clusters, **R-square index** was used. It measures the share of between-groups variability SS_B on total SS_T variability, see equation (3). Generally, an increasing number of these clusters become more homogeneous, and therefore the value of the determination ratio increases. A similar index based on an analysis of variance is **Calinski Habarasz F index** (*CHF*) as given in equation (4). It holds true that high values of *CHF* show well-formed clusters. In the analysis, therefore, the maximum within a certain interval is searched. From Table 3 it is obvious that both indices reach the highest values in case of 9 clusters. On the other hand, it is evident that when there is a large number of clusters, low-element clusters are formed. They include outlying observations, and are difficult to interpret. Therefore, 6 clusters, which were subjected to a further analysis, are the most sensible number of clusters.

$$RSQ = \frac{SS_T - SS_W}{SS_T} \tag{3}$$

$$CHF = \frac{\frac{SS_B}{k-1}}{\frac{SS_W}{n-k}} \tag{4}$$

Cluster No. k	Total squares B	Total squares W	RSQ	CHF
4	5.68	39.75	0.13	10.06
5	9.11	39.45	0.19	12.12
6	15.65	34.79	0.31	18.80
7	24.57	33.83	0.42	25.17
8	38.72	31.12	0.55	36.80
9	49.93	14.51	0.77	88.58
10	65.37	27.70	0.70	53.74

Table 3 RSQ and CHF indexes for different number of clusters

Municipalities in **cluster 1** include small to medium-sized municipalities (the largest municipalities is Cvikov with 4.5 thousand inhabitants, the smallest Janovice with 80 inhabitants). The cluster is characterized by favorable employment features, domicile attractiveness, age structure and branch structure. By contrast, the factor of population settlement and civic amenities develops negatively (see Figure 3). Cluster 1 can be classified as rural, fulfilling mainly the function of housing and recreation.

Cluster 2 contains all four district towns, all 10 seats of MEP and 5 other larger communities. In this cluster, population settlement, age structure, civic amenities and the branch structure can be positively evaluated. On the contrary, factors of unemployment, the domicile attractiveness and economic activity appear to be problematic. Cluster 2 can be described as urban, providing catchment function for the surrounding communities and branch focus on service.

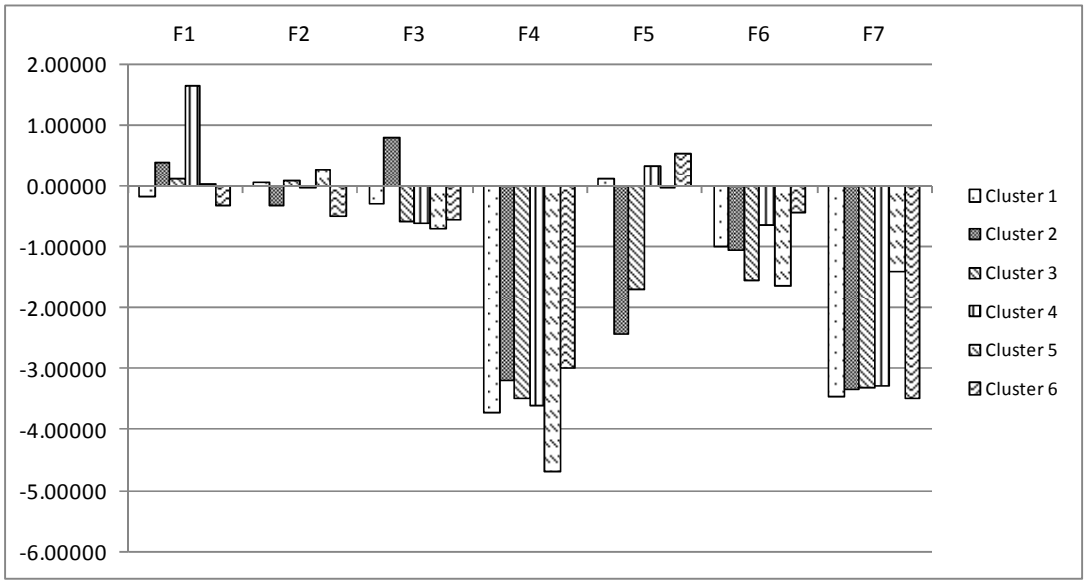


Figure 3 Average values of factor scores in clusters 1 to 6

In **cluster 3**, there are medium-sized municipalities (the average population of around 2 thousand), out of which 7 of them are municipalities with authorized municipal office that perform the functions of government. Other communities are centres of subregional units. The cluster is characterized by good domicile attractiveness and convenient civic amenities. The weaknesses are unemployment and population settlement. The cluster can be described as micro-regional providing essential central functions.

Cluster 4 includes rather medium-sized municipalities (average population of about 800), which are located along the borders of the Liberec Region. Of all the observed clusters, this one is characterized by high unemployment, negative domicile attraction, sparse population settlement and poor age structure. The cluster can, therefore, be classified as economically weak area of the region.

Cluster 5 contains rather smaller municipalities (except Plavy) mostly in the northwestern part of the region. The cluster shows good characteristics in the domicile attractiveness, very good age structure and average ratings in civic facilities (mainly thanks to the presence of primary schools in these municipalities). The unemployment factor can be evaluated as neutral. The weakness of municipalities in the cluster is a low density of settlement. In the cluster, there are tourist attractive municipalities (Bezdez, Zahradky, Bedrichov). The cluster can be characterized as rural with a predominant focus on tourism services.

In the last **cluster 6**, there are mostly small communities in the southeast region (MEP Turnov, Semily and Zelezny Brod). The cluster is characterized by favorable employment parameters and age structure. On the contrary, more domicile attractiveness, settlement and civic facilities can be viewed as weakness. The branch structure shows a higher proportion of a primary sector. The cluster can be classified as rural, with a higher share of employment in agriculture.

4 Conclusion

Although the determination of "optimal" number of clusters is relatively complex and ambiguous question, the analysis provides quite logical results. If we split the region just into two clusters, in the first cluster, there are larger cities and in the latter, there are small, rural municipalities. In the case of three clusters, small municipalities can be further divided into municipalities focusing on agriculture and cross-border and submountaneous communities, where services play a greater role in tourism. When using k-means method the division of the region into 6 clusters was eventually chosen. They can be classified as urban, rural - with the function of housing, rural - focused on services, rural - agricultural, micro-regional with central functions and the economically weak area. This division can be used as appropriate in formulating suitable development strategies of cities, municipalities and microregions. The creation of more clusters does no longer seem reasonable, since further created clusters begin to contain fewer elements.

Acknowledgements

The article was prepared with the support of the project Technology Agency of the Czech Republic ev. No. TD010029 "Definition of subregions for distinguishing and the solution of social and economic disparities".

References

[1] Antlová, K., Popelínský, L., and Tandler, J.: Long Term Growth of SME from ICT Competencies and Web Presentation. *E&M Ekonomie a Management* **14** (2011), 125-139.
[2] Draessler, J., Soukal, I., and Hedvičáková, M.: Shluková analýza poptávkové strany trhu základních bankovních služeb. *E&M Ekonomie a Management* **14** (2011), 102-114.
[3] Hebák, P., et al.: *Vícerozměrné statistické metody [3].* INFORMATORIUM, Prague, 2007.
[4] Kraftová, I., and Kraft, J.: High tech firmy a tvorba bohatství v zemích EMEA. *E&M Ekonomie a Management* **11** (2008), 6-20.
[5] Meloun, M., and Militký, J.: *Kompendium statistického zpracování dat.* Academia, Prague, 2006.
[6] Odehnal, J., and Michálek, J.: Empirická analýza regionálního podnikatelského prostředí vybraných zemí EU. *Politická ekonomie* **59** (2011), 242-262.
[7] Řezanková, H.: Hodnocení kvality shluků. In: *Analýza dat 2008/II – Statistické metody pro technologii a výzkum* (Kupka, K., ed.). TriloByte Statistical Software, Pardubice, 2008, 19-40.
[8] Žižka, M., et al.: Atlas ekonomických ukazatelů. Professional Publishing, Prague, 2011.
[9] 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.

Cumulative prospect theory and almost stochastic dominance in valuation of decision alternatives

Renata Dudzińska-Baryła[1], Ewa Michalska[2]

Abstract. Searching for good decision rules is one of the most important direction of research in decision making. There are well known and objective rules consistent with rationality, for example stochastic dominance rules, but, according to research based on behavioural approach, decision makers don't always act rationally. Relatively new tools, which model real choices, are cumulative prospect theory rules and almost stochastic dominance rules. The aim of our paper is to examine the consistency of the valuation of decision alternatives based on the cumulative prospect theory and the almost stochastic dominance rules. We show that choices made on the basis of considered tools are not always consistent, but the identification of causes of this discrepancy needs further research.

Keywords: cumulative prospect theory, almost stochastic dominance, stochastic dominance, decision making.

JEL Classification: D81
AMS Classification: 91B06

1 Introduction

For years the most researchers of the decision theory try to find new tools which will better model real decision-makers' choices. The prospect theory based on behavioral economics was one of such tools. Authors of the prospect theory were accused of its inconsistency with the stochastic dominance rules. That problem was solved by the cumulative prospect theory [13]. On the other hand also the stochastic dominance rules do not settle decision situations which would appear obvious. It forced some relaxation of those rules in the form of the almost stochastic dominance rules [6].

The aim of our paper is to examine the consistency of valuation of decision alternatives based on cumulative prospect theory and almost stochastic dominance rules.

2 Cumulative prospect theory

The prospect theory is one of the decision theories which try to explain the way decision-makers make their decisions in the situations of risk. In cumulative prospect theory [5, 13] the phase of evaluation of random decision alternative (prospect) is preceded by the editing phase. The aim of the editing phase is to organize and reformulate the prospects. Possible outcomes of prospect are transformed into gains and losses relative to some reference point which can represent the desirable or actual level of wealth. Then the representation of random decision alternative is different from that in the expected utility theory, in which the absolute levels of wealth are considered. Moreover, in the editing phase probabilities associated with the same outcomes are aggregated what simplifies further evaluation. As a result of the editing phase we obtain the prospect L represented as a sequence of relative outcomes x_i and corresponding probabilities p_i

$$\mathrm{L} = \left((x_1, p_1); \ldots; (x_k, p_k); (x_{k+1}, p_{k+1}); \ldots; (x_n, p_n)\right)$$

where $x_1 < \ldots < x_k < 0 \leq x_{k+1} < \ldots < x_n$ and $p_1 + \ldots + p_k + p_{k+1} \ldots + p_n = 1$.

In the second phase (evaluation phase) the value of each prospect is calculated. This value depends on two functions: value function $v(x)$ and probability weighting function $g(p)$. The analytical form of the value function and the evaluation of its parameters are determined on the basis of revealed preferences of decision-makers. In the literature various examples of the value function can be found (Dudzińska-Baryła and Kopańska-Bródka proposed quadratic function [2]) but the most cited is

[1] University of Economics in Katowice, Department of Operations Research, ul. 1 Maja 50, 40-287 Katowice, Poland, renata.dudzinska@ue.katowice.pl.
[2] University of Economics in Katowice, Department of Operations Research, ul. 1 Maja 50, 40-287 Katowice, Poland, ewa.michalska@ue.katowice.pl.

$$v(x) = \begin{cases} -\lambda(-x)^{\beta}, & x < 0 \\ x^{\alpha}, & x \geq 0 \end{cases}$$

in which values of parameters α, β and λ are 0.88, 0.88 and 2.25 respectively [13].

The plot of the value function is showed in figure 1. The S-shape of the value function is justified by the fact that losses make decision-maker risk-prone but when faced gains decision-maker is risk-averse.

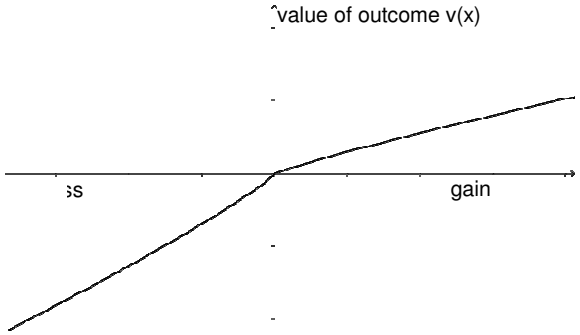


Figure 1 The value function

In cumulative prospect theory there is also considered the fact that decision-makers do not follow the objective probabilities. Therefore Tversky and Kahneman have proposed non-linear transformation of probabilities in the following probability weighting function

$$g(p) = \frac{p^{\gamma}}{\left[p^{\gamma} + (1-p)^{\gamma}\right]^{1/\gamma}}$$

where the value of parameter γ depends on that whether probability concerns gains or losses. In the literature [13] γ is 0.61 for gains and γ is 0.69 for losses. The plot of the probability weighting function is displayed in figure 2.

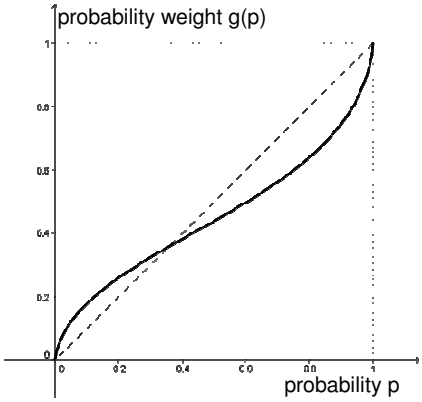


Figure 2 The probability weighting function

The probability weighting function has to have following features: it is increasing function, it overweights low probabilities, and underweights moderate and high probabilities, moreover $g(0) = 0$, $g(1) = 1$ and $g(p) + g(1-p) < 1$ for all $p \in (0,1)$. In the literature various probability weighting functions are analyzed, e.g. Currim and Sarin [1] have proposed four different forms of $g(p)$, and Prelec [12] and Wu and Gonzalez [14, 3] examined features of these functions.

Based on $v(x)$ and $g(p)$ functions the measure of the prospect value is constructed, which is the sum of the evaluation of gains $\mathrm{CPT}^{+}(\mathbf{x},\mathbf{p})$ and the evaluation of losses $\mathrm{CPT}^{-}(\mathbf{x},\mathbf{p})$ [13]:

$$\mathrm{CPT}(\mathbf{x},\mathbf{p}) = \mathrm{CPT}^{+}(\mathbf{x},\mathbf{p}) + \mathrm{CPT}^{-}(\mathbf{x},\mathbf{p})$$

Components $\mathrm{CPT}^{+}(\mathbf{x},\mathbf{p})$ and $\mathrm{CPT}^{-}(\mathbf{x},\mathbf{p})$ are calculated as follows:

$$\mathrm{CPT}^{+}(\mathbf{x},\mathbf{p}) = v(x_n)g(p_n) + \sum_{i=k+1}^{n-1} v(x_i)\left[g\left(\sum_{j=i}^{n} p_j\right) - g\left(\sum_{j=i+1}^{n} p_j\right)\right]$$

$$\mathrm{CPT}^{-}(\mathbf{x},\mathbf{p}) = v(x_1)g(p_1) + \sum_{i=2}^{k} v(x_i)\left[g\left(\sum_{j=1}^{i} p_j\right) - g\left(\sum_{j=1}^{i-1} p_j\right)\right]$$

In the evaluation phase for each prospect the measure $\mathrm{CPT}(\mathbf{x},\mathbf{p})$ is calculated. Among all prospects one with the highest value is preferred. The dominance rule based on cumulative prospect theory can be formulated as follows:

CPT: Alternative L1 dominates alternative L2 (written as $\mathrm{L1} \succ_{\mathrm{CPT}} \mathrm{L2}$) if and only if $\mathrm{CPT}(\mathrm{L1}) > \mathrm{CPT}(\mathrm{L2})$.

3 Almost stochastic dominance rules

For years the most common decision rule under risk was the mean-variance (MV) rule proposed by Markowitz [10]. For risky alternative $\mathrm{L1}$ and $\mathrm{L2}$ with expected values $\mathrm{E(L1)}$, $\mathrm{E(L2)}$ and standard deviations $\sigma(\mathrm{L1})$, $\sigma(\mathrm{L2})$ the MV rule is following:

MV: Alternative L1 dominates alternative L2 (written as $\mathrm{L1} \succ_{\mathrm{MV}} \mathrm{L2}$) if and only if $\mathrm{E(L1)} \geq \mathrm{E(L2)}$ and $\sigma(\mathrm{L1}) \leq \sigma(\mathrm{L2})$ with at least one strict inequality.

Common accepted and objective nonparametric decision rule is the stochastic dominance. Lets $\mathrm{F_{L1}}$ and $\mathrm{F_{L2}}$ be the distribution functions of risky alternative $\mathrm{L1}$ and $\mathrm{L2}$ respectively, and S be a set of all outcomes of $\mathrm{L1}$ and $\mathrm{L2}$. The first and the second stochastic dominance rules are formulated as follows [4]:

FSD: Alternative $\mathrm{L1}$ dominates alternative $\mathrm{L2}$ by the first stochastic dominance (written as $\mathrm{L1} \succ_{\mathrm{FSD}} \mathrm{L2}$) if and only if inequality $\mathrm{F_{L1}}(r) - \mathrm{F_{L2}}(r) \leq 0$ is satisfied for each $r \in \mathrm{S}$ and for at least one value $r \in \mathrm{S}$ this inequality is strict.

SSD: Alternative $\mathrm{L1}$ dominates alternative $\mathrm{L2}$ by the second stochastic dominance (written as $\mathrm{L1} \succ_{\mathrm{SSD}} \mathrm{L2}$) if and only if inequality $\mathrm{F_{L1}^{(2)}}(r) - \mathrm{F_{L2}^{(2)}}(r) \leq 0$ is satisfied for each $r \in \mathrm{S}$ and for at least one value $r \in \mathrm{S}$ this inequality is strict, where $\mathrm{F_{L1}^{(2)}}(r) = \int_{-\infty}^{r} \mathrm{F_{L1}}(t)\,dt$ and $\mathrm{F_{L2}^{(2)}}(r) = \int_{-\infty}^{r} \mathrm{F_{L2}}(t)\,dt$.[3]

The MV and stochastic dominance rules often do not lead to the conclusion which alternative is better. In such situation we need other criteria for decision-making. Such situation is presented in example 1.

Example 1. The possible results of the risky alternative $\mathrm{L1}$ are to gain \$1 with probability 0.01 or to gain \$100 with probability 0.99, and in the alternative $\mathrm{L2}$ one can gain certain \$2. Both alternatives can be written as $\mathrm{L1} = ((1, 0.01); (100, 0.99))$ and $\mathrm{L2} = ((2, 1))$. It is easy to show that neither $\mathrm{L1}$ dominates $\mathrm{L2}$ nor $\mathrm{L2}$ dominates $\mathrm{L1}$ based on the MV rule. Also neither $\mathrm{L1}$ nor $\mathrm{L2}$ dominates the other based on the first or the second stochastic dominance rules, but most "reasonable" decision-makers (if not all) prefer $\mathrm{L1}$ to $\mathrm{L2}$. Moreover, analyzing graphs of both distribution functions showed in figure 3, we can notice that the area A corresponding to the range in which $\mathrm{L2}$ dominates $\mathrm{L1}$, is much smaller than the area B corresponding to the range in which $\mathrm{L1}$ dominates $\mathrm{L2}$. Therefore we can say that $\mathrm{L1}$ "almost" dominates $\mathrm{L2}$ by the first stochastic dominance.

[3] For discrete probability distributions the values $\mathrm{F_{L1}^{(2)}}(r)$ and $\mathrm{F_{L2}^{(2)}}(r)$ are cumulated values of the distribution functions (sums of the cumulated probabilities).

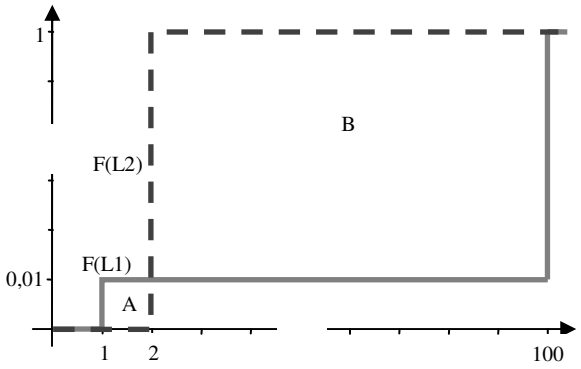


Figure 3 The distribution functions for L1 and L2

Analyzing similar examples Leshno and Levy proposed the concept of almost stochastic dominance (ASD) which is some relaxation of stochastic dominance rule [6]. The definitions of almost first and second stochastic dominance are as follows:

AFSD: Alternative L1 dominates alternative L2 by almost first stochastic dominance (written as $L1 \succ_{AFSD} L2$) if and only if

$$\int_{S_1} (F_{L1}(r) - F_{L2}(r))\, dr \le \varepsilon \int_{S} |F_{L1}(r) - F_{L2}(r)|\, dr$$

where S is a set of all outcomes of L1 and L2 and $S_1 = \{r \in S : F_{L2}(r) < F_{L1}(r)\}$.

ASSD: Alternative L1 dominates alternative L2 by almost second stochastic dominance (written as $L1 \succ_{ASSD} L2$) if and only if

$$\int_{S_2} (F_{L1}(r) - F_{L2}(r))\, dr \le \varepsilon \int_{S} |F_{L1}(r) - F_{L2}(r)|\, dr$$

and

$$E(L1) \ge E(L2)$$

where S is a set of all outcomes of L1 and L2 and $S_2 = \{r \in S_1 : F^{(2)}_{L2}(r) < F^{(2)}_{L1}(r)\}$.

It is assumed that the value of ε parameter connected with "actual" violation area should be less than 0.5 for both the first and the second stochastic dominance rules. [4]

In the example 1 neither alternative L1 nor alternative L2 dominates the other, and second stochastic dominance, but L1 dominates L2 by AFSD for $\varepsilon \approx 0.000103$ (parameter ε is defined as the area A divided by the total absolute area enclosed between both distribution functions (area A+B)).The main advantage of applying almost stochastic dominance rules is the possibility for reduction of a set of non-comparable (according to other criteria) risky alternatives. Moreover, almost stochastic dominance rules reveal preferences consistent with intuition, whereas traditional stochastic dominance rules may not confirm intuitional choices.

4 Consistency between preferences based on the cumulative prospect theory and the almost stochastic dominance rules

To analyze the consistency between preferences determined on the basis of the cumulative prospect theory and the almost stochastic dominance rules we have examined some examples of pairs of decision alternatives.

For alternatives L1 and L2 (showed in example 1) the selection of dominating alternative on the basis of the cumulative prospect theory is consistent with the selection based on almost stochastic dominance rules (sum-

[4] In literature [8] there are also defined ε^*−AFSD and ε^*−ASSD, where ε^* indicates "allowed" violation area, and $0 < \varepsilon < \varepsilon^* < 0.5$.

mary of calculated values is in table 1 and 2). The question arises whether this consistency will always be observed?

Example 2. Let's consider two risky alternative $L3 = ((30, 0.4);(60, 0.6))$ and $L4 = ((40, 0.4);(50, 0.6))$. Similarly as for the pair $L1$ and $L2$ none of the alternative dominates the other by the MV rule (see table 1 and 2). There is also no dominance by FSD rule (what can be seen in figure 4) and SSD rule. But selections made by CPT and AFSD (and consequently ASSD) rules do not coincide. According to CPT rule $L4$ is the dominating alternative and according to AFSD (and also ASSD) $L3$ is the dominating one (see table 1 and 2).

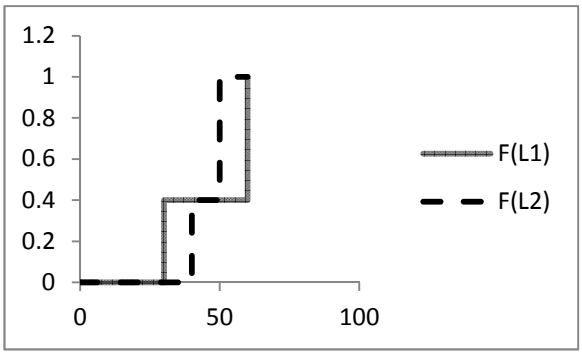


Figure 4 The distribution functions for $L3$ and $L4$

Example 3. Let's consider alternatives $L5 = ((20,0.2);(30,0.5);(56,0.3))$ and $L6 = ((10,0.1);(28,0.5);(52,0.4))$. The alternative $L5$ dominates $L6$ by MV rule, but there is no dominance by FSD, SSD and AFSD (see table 1 and 2). It is worth to notice that selections based on the CPT and ASSD do not coincide. Alternative $L5$ is dominating according to CPT rule and alternative $L6$ is dominating by ASSD. Our example is also interesting because both ε (for ASSD) are less than 0,5 and both expected values are the same. In such case the alternative with lower epsilon dominates [6]. Therefore $L6$ dominates $L5$ by ASSD rule.

	Alternatives					
Parameter	L1	L2	L3	L4	L5	L6
E(L)	99.01	2.00	48.00	46.00	35.80	35.80
σ(L)	9.85	0.00	14.70	4.90	13.75	14.21
CPT(L)	52.54	1.84	27.89	28.33	22.24	20.58
ε_{AFSD}	0.000103 for $(L1,L2)$		0.4 for $(L3,L4)$		0.5 for $(L5,L6)$	
	0.999897 for $(L2,L1)$		0.6 for $(L4,L3)$		0.5 for $(L6,L5)$	
ε_{ASSD}	0.000103 for $(L1,L2)$		0.4 for $(L3,L4)$		0.366667 for $(L5,L6)$	
	0.999897 for $(L2,L1)$		0.6 for $(L4,L3)$		0.3 for $(L6,L5)$	

Table 1 Values of parameters for alternatives $L1-L6$

Decision rule	Example 1	Example 2	Example 3
MV	-	-	$L5 \succ L6$
CPT	$L1 \succ L2$	$L4 \succ L3$	$L5 \succ L6$
FSD	-	-	-
SSD	-	-	-
AFSD	$L1 \succ L2$	$L3 \succ L4$	-
ASSD	$L1 \succ L2$	$L3 \succ L4$	$L6 \succ L5$

Table 2 Dominances for alternatives $L1-L6$

Mutual dependences between considered decision rules are showed on diagram in figure 5.

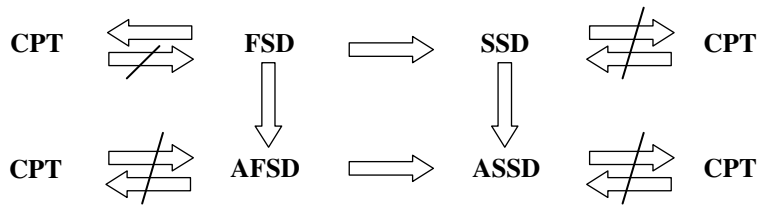


Figure 5 Dependences between FSD, SSD, AFSD, ASSD and CPT rules

The dependences showed in figure 5 between FSD, SSD, AFSD and ASSD dominances are supported by the literature [6, 7]. In the analyzed examples we showed that there is no consistency between choices based on cumulative prospect theory and the almost first and second stochastic dominance. Correctness of implication FSD $\Rightarrow$ CPT is confirmed in [9, 13], whereas converse implication do not occur [11]. Some examples showing no consistency between CPT and SSD rules are presented in [11].

5 Summary

The motivation of the authors of the cumulative prospect theory as well as the almost stochastic dominance rules was to create good tools for modeling real choices. However, as we showed in our article choices made on the basis of considered tools are not always consistent. The identification of causes of this discrepancy will need further research.

References

[1] Currim I., Sarin R.: Prospect versus utility. *Management Science* **35** (1989), No. 1, 22–41.
[2] Dudzińska-Baryła R., Kopańska-Bródka D.: Maximum expected utility portfolios versus prospect theory approach. In: *Increasing competitiveness or regional, national and international markets development..Proceedings of the 25th International Conference on Mathematical Methods in Economics 2007*. Technical University of Ostrava, Ostrava, 2007, electronic document.
[3] Gonzalez R., Wu G.: On the shape of the probability weighting function. *Cognitive Psychology* **38** (1999), 129–166.
[4] Hanoch G., Levy H.: The efficiency analysis of choices involving risk. *Review of Economic Studies* **36** (1969), No. 3, p. 335–346.
[5] Kahneman D., Tversky A.: Prospect Theory: An Analysis of Decision Under Risk. *Econometrica* **47** (1979), 263–291.
[6] Leshno M., Levy H.: Preferred by "all" and preferred by "most" decision makers: Almost stochastic dominance. *Management Science* **48** (2002), Issue 8, 1074–1085.
[7] Levy H.: Stochastic dominance and expected utility: survey and analysis. *Management Science* **38** (1992), No. 4, 555–593.
[8] Levy H., Leshno M., Leibovitch B.: Economically relevant preferences for all observed epsilon. *Annals of Operations Research* **176** (2010), 153–178.
[9] Levy H., Wiener Z.: Prospect theory and utility theory: temporary versus permanent attitude towards risk. *http://pluto.mscc.huji.ac.il/~mswiener/research/PTandUT.pdf* , 2002
[10] Markowitz H.: Portfolio selection. *The Journal of Finance* **7** (1952), No. 1, 77–91.
[11] Michalska E.: Dominacje stochastyczne a teoria perspektyw. In: *Współczesne tendencje rozwojowe badań operacyjnych,* prace naukowe nr 108 (Siedlecki, J., Peternek, P., eds.), Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław, 2010, 167–176.
[12] Prelec D.: The probability weighting function. *Econometrica* **66** (1998), No. 3, 497–527.
[13] Tversky A., Kahneman D.: Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty* **5** (1992), 297–323.
[14] Wu G., Gonzalez R.: Curvature of the probability weighting function. *Management Science* **42** (1996), 1676–1690.